Optimizing Athlete Training Load and Recovery Monitoring in Hockey

KONERTH, NATALIE,MARIE

Optimizing Athlete Training Load and Recovery Monitoring in Hockey

Natalie Konerth

Abstract

When implemented correctly, athlete monitoring can be used to elevate performance, decrease injuries, and improve wellbeing in athletes at all levels. The demands of hockey are distinct from other intermittent ball sports, and monitoring is currently limited in that the metrics monitored have not been optimized for use in hockey. Therefore, this research program aimed to produce an evidence-based model for athlete monitoring in hockey, with four subsequent objectives addressed via four distinct studies on various components of monitoring: internal training load, external training load and recovery. The match-demands of hockey competition were first evaluated, via a systematic review and metanalysis, and elite female and male athletes were found to cover 5029 ± 424 m and 6027 ± 536 m in competition, respectively, with an average workrate of 115 ± 8 $m \cdot min^{-1}$ and 125 ± 8 $m \cdot min^{-1}$. To evaluate the validity of recovery monitoring measures, the relationship between training load and recovery was evaluated. Countermovement jump height was shown to have no substantial association with training load ($r = -0.06, -0.09$, $p = 0.506- 0.568$), but the Recovery-Stress Questionnaire for Athletes general and sports stress subscales were responsive to changes in load ($r = 0.47 – 0.57$, $p = 0.006 – 0.030$). External training load was evaluated via the validity and reliability of Catapult S7 Global Navigation Satellite System units, which were shown to have a small mean negative bias of 2.8%, with good reliability (%SEE: 0.98%). Finally, a new pitch-based testing protocol and algorithm were developed for the calculation of internal training load (piTRIMP2), which outperformed existing metrics, explaining 84% of the variability in athlete fitness over a hockey season. The results of these studies were used to develop a novel evidence-based model for athlete monitoring in hockey which provides a framework for implementing athlete monitoring systems across hockey populations.

Optimizing Athlete Training Load and Recovery Monitoring in Hockey

Natalie Konerth

A thesis submitted for the degree of Doctor of Philosophy

Department of Sport and Exercise Sciences

Durham University

2023

Table of Contents

Abbreviations

% Max HR	Percentage of maximum heart rate
%SEE	Percent standard error of the estimate
30-15 IFT	30-15 intermittent fitness test
ANS	Autonomic nervous system
ARSS	Acute Recovery and Stress Scale
AU	Arbitrary units
BLvHR	Blood lactate versus heart rate
CI	Confidence interval
CK	Creatine kinase
CMJ	Countermovement jump
COD	Change of direction
CV	Coefficient of variation
DALDA	Daily Analyses of Life Demands for Athletes
EWMA	Exponentially weighted moving average
fTRIMP	Female training impulse
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HRA	Heart rate acceleration
HRR	Heart rate recovery
HRV	Heart rate variability
HSR	High speed running
ICC	Interclass correlation coefficient
iTRIMP	Individualized training impulse
iTRIMP1	Laboratory-based individualized training impulse calculated with the heart rate reserve term
iTRIMP2	Laboratory-based individualized training impulse calculated the secondary heart rate reserve term
$LnRMSSD_{CV}$	coefficient of variation of the logarithm of the root mean square successive differences
piTRIMP	Pitch-based individualized training impulse
piTRIMP1	Pitch-based individualized training impulse calculated with the heart rate reserve term
piTRIMP2	Pitch-based individualized training impulse calculated the secondary heart rate reserve term
POMS	Profile of Mood Scores
RESTQ-S	Recovery-Stress Questionnaire for Athletes
RMSSD	root mean square of differences of successive RR intervals

RPE	Rating of percieved exertion
sRPE	Session rating of percieved exertion
SRSS	Short Recovery and Stress Scale
sTRIMP	Stagno's training impulse
TD	Total distance
TE	Typical error
TRIMP	Training impulse
v_{OBLA}	Velocity at the onset of blood lactate accumulation

Declaration

The work of this thesis is based on the research carried out by Natalie Konerth under the supervision of Rob Cramb, Caroline Dodd-Reynolds, and Karen Hind, within the Department of Sport and Exercise Science, Durham University, United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification.

Statement of Copyright

Acknowledgements

I would like to thank my supervisors, Rob Cramb, Caroline Dodd-Reynolds, and Karen Hind for their guidance and effort throughout my research. Special thank you to Rob for being there from start to finish and supporting me to achieve my dream. I am forever grateful.

This work was only achieved through the unceasing support of Durham University Hockey Club. Thank you for welcoming me into your family and to all the athletes and coaches who participated in and supported my research. Special thank you to Gareth Weaver-Tyler as coach, mentor, and friend on this journey. None of this would have been possible without you.

To the Konerth Clan,

Chapter 1: Introduction

1.1 Introduction to hockey

Hockey is the world's oldest stick and ball game, with origins dating back to Asia in approximately 2000 BC (Reilly and Seaton, 1990; Lythe, 2008). The sport has evolved over the millennia with the first hockey club formed in 1861 and the first official international matches also taking place in the late 19th century (Reilly and Borrie, 1992; Lythe, 2008). Making its Olympic debut in 1908, hockey is now an Olympic sport for both men and women (Olympic.org, 2023). Hockey is governed by the International Hockey Federation (FIH) which recognizes 139 national associations across 6 continents (International Hockey Federation, 2023b). There are 80 women's national teams and 96 men's national teams ranked by the FIH, indicating that several thousand athletes participate in hockey at the senior international level (International Hockey Federation, 2023a). Club hockey is also incredibly popular, with professional and semi-professional leagues taking place annually within several continents (Barboza *et al.*, 2018). In the England alone, over 140,000 athletes participate in club hockey, with approximately 15,000 playing for their college/university (England Hockey, 2023).

Despite the popularity of the sport, there is a dearth of academic literature on hockey in comparison to other more lucrative field-based team sports, such as football and rugby (Podgórski and Pawlak, 2011; Drew and Finch, 2016; Eckard *et al.*, 2018; Fox *et al.*, 2018; McLaren *et al.*, 2018). Specifically, although a range of studies have described the physical and physiological demands of hockey competition, there is no consensus on the best variables and protocols for monitoring hockey athletes (Konerth, 2019). As shown in studies of other intermittent ball sports, individualized athlete monitoring has many potential benefits, such as reducing overtraining and improving performance (Meeusen *et al.*, 2013; Drew and Finch, 2016; Coutts, Crowcroft and Kempton, 2017; Eckard *et al.*, 2018; West *et al.*, 2020). However, without appropriate research on the most effective methods and evidence-based variables for monitoring training dose and evaluating athlete recovery status, these benefits cannot be realized (Gabbett *et al.*, 2017; Thornton *et al.*, 2019).

1.2 Training load and recovery monitoring

Athletic training is the process of performing exercise to elicit a physical adaptation or acquisition of sport skills (Impellizzeri, Marcora and Coutts, 2022). Thus, the goal of exercise in this scenario is to create a psychophysiological response that becomes the stimulus for positive adaption (Coutts, Crowcroft and Kempton, 2017; Impellizzeri, Marcora and Coutts, 2022). However, the nature of the response is specific to the type of exercise and is dependent on the timing, intensity, and composition of the exercise (Impellizzeri, Marcora and Coutts, 2022). Functional adaptation in athletic training is also dependent on the appropriate exercise being performed at the correct time and load prior to competition (Jeffries *et al.*, 2020). To promote positive adaption, athlete monitoring can be used to measure and quantify the nature of the exercise to refine the training process (Coutts, Crowcroft and Kempton, 2017).

Training load is a measure of an athlete's work during a training session or competition, summarized as a single numerical score (Impellizzeri, Marcora and Coutts, 2022). Training load measures can be internal, monitoring the physiological demands of exercise on the body, or external, focusing solely on physical output regardless of the physiological response (Impellizzeri, Marcora and Coutts, 2022). Early forms of training load monitoring began in endurance runners with training logs used to track external load in terms of total distance (Foster *et al.*, 2001). However, without information on the speed at which the distance was covered and the related physiological demands, the benefits of these training logs were limited (Foster *et al.*, 2001). Furthermore, in field-sports it is impossible to determine the distance covered by individual athletes without tracking devices or sophisticated camera setups, causing coaches to often erroneously rely on their intuition when making decisions regarding training dose (Bompa, 1999; Brink *et al.*, 2014; Kraft *et al.*, 2018). However, as technology has evolved, new more-sophisticated methods of measuring training load have emerged, and there has been a subsequent increase in training load monitoring in team-sport athletes (Cummins *et al.*, 2013; Drew and Finch, 2016; Eckard *et al.*, 2018; Torres-Ronda *et al.*, 2022). Athlete monitoring is of particular importance in intermittent sport athletes because the change of speed and direction, and corresponding acceleration and deceleration, impact the manner in which training load is accumulated (Dellal *et al.*, 2010; Harper, Carling and Kiely, 2019; Bekraoui *et al.*, 2020). As a result, training load is far more difficult to monitor and measure than in straight line running.

Recovery monitoring measures athletes' overall wellbeing, determining how they are responding to a given training dose outside of a sport-specific setting (Saw, Main and Gastin, 2016). Whereas training load is measured during training, recovery monitoring takes into consideration time spent outside of sport, with athletes' actions, decisions, and stress levels impacting their physiological response to training (Sperlich and Holmberg, 2017). For optimal performance and wellbeing, athletes must maintain a balance between stress and recovery (Kallus and Kellmann, 2016). When under-recovery occurs, athletes are at risk of developing maladaptive training states of non-functional overreaching and overtraining, which are detrimental to both performance and health (Meeusen *et al.*, 2013). As individual athletes respond differently to identical training doses, regular individualized recovery monitoring is key to ensuring that athletes are adapting positively (Bourdon, 2017). Recovery monitoring can be either subjective, often in the form of questionnaires, or objective, such as blood-based markers or measures of neuromuscular fatigue (for example, creatine kinase and countermovement jump height, respectively) (Saw, Main and Gastin, 2016). Although recovery monitoring is increasing in prevalence across sports, there is to date no gold-standard approach for measurement (Meeusen *et al.*, 2013; Bourdon, 2017).

1.3 Athlete monitoring in hockey

Despite the positive impact of athlete monitoring across other sports (Coutts, Crowcroft and Kempton, 2017; Gabbett *et al.*, 2017; Thornton *et al.*, 2019; West *et al.*, 2020), there is no consensus on the best methods for measuring training load or recovery in hockey athletes. Without these measures, it is not possible to develop an evidence-based model and optimize athlete monitoring in hockey. Although hockey is similar in many ways to other intermittent ball sports such as football, rugby, and lacrosse, it has several significant distinctions that make it unique, illustrating the need for sport-specific research (Reilly and Seaton, 1990; White and MacFarlane, 2013; Abbott, 2016; McGuinness *et al.*, 2017). Furthermore, the research which directly preceded this thesis showed that the demands of hockey are highly variable, making individualized monitoring of particular importance in hockey athletes (Konerth, 2019).

Unlike many other intermittent ball sports, hockey has unlimited rolling substitutions, meaning that athletes on the pitch can interchange with the 5-7 athletes on the bench at almost any time (Abbott, 2016). As a result, hockey is played at a higher intensity than other field-based team sports, with athletes averaging 85-89% of their maximum heart rate while on the pitch (Lythe, 2008; Sell and Ledesma, 2016; Vescovi, 2016; McGuinness *et al.*, 2017). Rolling substitutions also provide an increased challenge for training load measurement and analysis, as time spent on the bench confounds time-dependent measures such as average speed, if data are not individually phased (White and MacFarlane, 2013). Additionally, athletes' recovery time on the bench, during which heart rate may still be elevated, impacts physiological load measures (Konerth, 2019). Hockey has no offsides or restraining lines, causing athlete movement patterns to be stochastic in nature and impossible to accurately predict (McGuinness *et al.*, 2017). Finally, hockey athletes assume a semi-crouched posture, distinct from other sports, while passing, receiving, dribbling, and defending (Reilly and Seaton, 1990). This posture has been shown to increase perceived exertion and energy expenditure (additional 15-16 $kJ \cdot min^{-1}$) compared to normal running (Reilly and Seaton, 1990). As a result of these differences, it is almost impossible to accurately apply studies performed on other intermittent ball sports to hockey, demonstrating the need for research specific to this population.

The demands of hockey are extremely varied, even within teams, highlighting the importance of athlete monitoring in hockey (Gabbett, 2010; Boran, 2012; Sunderland and Edwards, 2017; Vescovi, 2016; McGuinness *et al.*, 2017; Konerth, 2019). Specifically, the physical demands of the game vary by playing position, with average differences of up to 2.3 km in total distance per match (Gabbett, 2010; Boran, 2012; Sunderland and Edwards, 2017; Vescovi, 2016; McGuinness *et al.*, 2017; Konerth, 2019). Although there are differences across teams, defenders generally play the most minutes and cover the greatest low speed distance, while forwards and attacking midfielders generally play for fewer minutes but cover substantially more distance at high speeds (Gabbett, 2010; Macutkiewicz and Sunderland, 2011; Boran, 2012; Jennings *et al.*, 2012c; McGuinness *et al.*, 2017). Without training load monitoring, coaches are unable to definitively determine the work performed by each athlete and relate this to their requirements for rest and recovery (Brink *et al.*, 2014). As a result, some athletes may become fatigued or overtrained and others undertrained, while participating in identical matches and training sessions (Drew and Finch, 2016; Eckard *et al.*, 2018).

As with other intermittent ball sports, athlete monitoring has the potential to improve performance, minimize injuries, and increase athlete wellbeing in hockey. However, these benefits are currently limited by the lack of research on the best methods for measuring training load and recovery in hockey populations. Additionally, the match-demands of hockey have not been comprehensively reviewed. Therefore, establishing the match-demands of hockey and developing an evidence-based model for athlete monitoring in hockey will allow for optimization of the athlete monitoring process, maximizing benefits to the athletes and improving the sport.

1.4 Impact of athlete monitoring

When performed appropriately, athlete monitoring can help coaches to improve athlete performance, minimize injuries, and increase wellbeing (Meeusen *et al.*, 2013; Drew and Finch, 2016; Coutts, Crowcroft and Kempton, 2017; Eckard *et al.*, 2018; West *et al.*, 2020).. As the most effective form of training is that which best mirrors the physical and physiological demands of competition, training load monitoring is also important in ensuring that athletes are best prepared for competition (Gabbett, 2010). Training programs that use periodization to regulate training dose and combine periods of high intensity and low intensity have been repeatedly shown to significantly improve performance and minimize overtraining (Morton, Fitz-Clarke and Banister, 1990; Busso *et al.*, 1997; Mujika, 1998; Bompa, 1999; Foster *et al.*, 2001; Meeusen *et al.*, 2013; Kevin and James, 2015; Mara *et al.*, 2015). However, without training load monitoring, it is impossible to know if training prescriptions are met and if appropriate periodization has been achieved. Furthermore, an athlete's recovery status is indicative of their overall wellbeing and sport performance (Bourdon, 2017). Recovery monitoring helps ensure that athletes maintain recovery-stress balance and avoid maladaptive training states (Meeusen *et al.*, 2013). Finally, although there is no causal relationship and injuries are outside the scope of this thesis, the results of several systematic reviews have demonstrated a relationship between athlete training load and injuries (Drew and Finch, 2016; Jones, Griffiths and Mellalieu, 2017; Eckard *et al.*, 2018) Therefore, athlete monitoring can be used to improve athletes' training to promote positive adaptation while avoiding injuries and overtraining, resulting in improved performance and wellbeing.

Given its many benefits when implemented correctly, athlete monitoring has become a key element of athlete management in elite sport (Coutts, Crowcroft and Kempton, 2017; Gabbett *et al.*, 2017; Thornton *et al.*, 2019; West *et al.*, 2020). Several frameworks for the training process have been developed, incorporating internal training load, external training load, and, in some cases, athlete wellbeing and fitness status, across other intermittent ball sports (Impellizzeri, Rampinini and Marcora, 2005; Coutts, Crowcroft and Kempton, 2017; Gabbett *et al.*, 2017; West *et al.*, 2020). Athlete monitoring systems can be used to collate and analyze these data to provide decision-support systems regarding an athlete's readiness status (Robertson, Bartlett and Gastin, 2017). This allows for the utilization of routinely collected training load and recovery monitoring data to inform decision-making, thereby maximizing the benefits of athlete monitoring (West *et al.*, 2020). However, an athlete monitoring system is only as good as its component measures of training load and recovery monitoring, with evidence-based variables and their relationships the foundation of any monitoring system (Thornton *et al.*, 2019). In order to develop an effective model for athlete monitoring and maximize the benefits of its implementation, evidence-based measures of training load and recovery are required.

1.5 Overview

Routine athlete monitoring can be broadly divided into two main categories – training load monitoring and recovery monitoring (Gabbett *et al.*, 2017). Training load monitoring considers the demands of exercise, most often occurring during training or competition sessions (Bourdon, 2017). When performed correctly, this form of athlete monitoring allows coaches and sport scientists to accurately measure both the acute and long-term physical and physiological demands on athletes (Bourdon, 2017). On the other hand, recovery monitoring evaluates how an athlete is responding to the physical, physiological, and even psychological demands of their sport, with this form of monitoring focusing on athletes' overall wellbeing and response to a training stimulus outside of their immediate sports environment (Duffield *et al.*, 2018) Training load monitoring considers the on-pitch, or otherwise sport-specific, demands on athletes, whereas recovery monitoring incorporates the impact of athletes' off-pitch behaviors on readiness to train (Sperlich and Holmberg, 2017). Thus, distinguishing athlete monitoring in this way allows for a macro view of athlete wellbeing status, separated based on training demands and the response to those demands. This separation also aligns with the recovery-stress model of athlete readiness, with training load representing stress on an athlete and recovery monitoring considering how athletes are able to respond to that stress (Kallus and Kellmann, 2016). With a balance of stress and recovery key in minimizing injuries, avoiding overtraining and improving performance, distinguishing measures in this way provides a natural framework for modeling athlete monitoring (Duffield *et al.*, 2018). Fitness testing is also considered an aspect of athlete monitoring systems; however, fitness is not monitored as frequently as load or recovery, which are typically measured on a sessional basis (West *et al.*, 2020). As such, athlete fitness may be considered distinct from other routine athlete monitoring measures, and it will not be investigated in the same detail here.

As outlined in Figure 1.1, this thesis will begin with an analysis of recovery monitoring and then consider training load, both external and internal. Chapter 3 provides a summary of the literature on athlete recovery monitoring, incorporating both subjective and objective measures. Internal and external training load in hockey are then reviewed via the first study in chapter 4, a systematic review and meta-analysis of match demands in competitive hockey. This study lays the foundation for research on athlete monitoring in hockey by establishing the demands of competition and considering the monitoring metrics currently in use. Two of these recovery metrics, the Recovery-Stress Questionnaire for Athletes and countermovement jump (CMJ) height, were then evaluated in chapter 5 alongside their relationship with athlete load to determine the potential validity of these metrics for monitoring recovery in hockey athletes. Chapters 6-8 consider internal and external training load measures in a hockey-specific context. Having established population values from the systematic review, chapter 6 evaluates the accuracy of external training load monitoring equipment, determining the validity and interunit reliability of Catapult Vector 10 Hz global navigation satellite system (GNSS) units for tracking hockey-specific athlete movement patterns. For internal training load, much of the discussion lies in the validity of the summary metrics and algorithm themselves rather than the technology. Thus, chapter 7 outlines the development of a new pitch-based protocol for calculating internal training load in intermittent sport athletes, and chapter 8 evaluates this new metric over the course of a hockey season. Bringing together both training load and recovery monitoring, chapter 9 presents an evidence-based model for athlete monitoring in hockey which can be implemented to improve monitoring, enhancing performance and athlete wellbeing.

Figure 1.1 Research Overview

1.6 Research aim and objectives

The overall aim of this research program was to develop an evidence-based model to optimize athlete monitoring (internal training load, external training load, and recovery monitoring) in hockey. Athlete monitoring is only as good as the monitoring techniques and metrics it consists of; therefore, the goal of this research was to determine the evidence-based measures for training load and recovery monitoring best suited for use in hockey.

In order to determine the ideal monitoring measures for hockey, the physical and physiological match-demands of hockey competition needed to be known, as well as the measures currently used for athlete monitoring in hockey populations. Therefore, the first objective of this research project was to determine the physical and physiological demands of hockey match-play. With this knowledge, it was then possible to assess the various aspects of athlete monitoring and evaluate the metrics best suited for use in hockey populations. Recovery monitoring was first evaluated to determine which recovery monitoring measures were most responsive to changes in training load in hockey populations. External training load was considered next; however, as summary measures for external load (such as distance and speed) are clearly established, the question of validity lay more in the monitoring devices themselves rather than the metrics. Therefore, the research objective was to evaluate the accuracy and precision of external training load monitoring using current GNSS models. Finally, internal training load monitoring was considered specifically in terms of training impulse to develop new protocols and algorithms for summarizing heart rate data into summary scores for hockey training and competition.

With the overall aim of developing an evidence-based model to optimize athlete monitoring in hockey, the objectives of this research program were as follows.

1. Determine the physical and physiological demands of hockey competition (Chapter 4).
2. Evaluate the relationship between athlete training load and recovery status in hockey (Chapter 5).
3. Assess the accuracy and precision of external training load monitoring in hockey using current GNSS models (Chapter 6).
4. Design an improved hockey-specific protocol and algorithm for calculating internal training load from heart rate data (Chapter 7 and 8).

The results from this research provide an applied framework for athlete monitoring in hockey, which can be implemented to improve athletes' performance, fitness, and overall wellbeing.

Chapter 2: Research Design and Methodology

2.1 Methodological approach

Before exploring athlete monitoring, it is first important to consider the philosophical underpinnings of this research, as determined by the epistemological and ontological approach. Whether implicit or explicit, the approach to any research questions is dependent one's philosophical assumptions (Scotland, 2012). Ontology, quite simply the study of being, considers what is known and exists (Scotland, 2012). Closely related, epistemology is the study of the nature and status of knowledge, considering how and what is known (Thomas, 2011). Therefore, before conducting any research, it is necessary to consider the ontological and epistemological assumptions that are being made and how these inform the research methods (Pisk, 2014).

Overall, there are two main philosophical approaches that can be adopted: positivism and interpretivism. Positivism has its foundations in the ontological position of realism, with an epistemological approach of objectivism (Scotland, 2012). This approach assumes that knowledge and reality exist independent of the researcher and are not influenced by our senses or language (Bryman, 1984). As such, the role of the researcher is to discover objective realities and impartial knowledge about the world (Thomas, 2011). Thus, positivists often seek to understand causal relationships from the outside using quantitative methods, with little consideration given to the context, beliefs or feelings of participants (Bryman, 1984). As positivists trust in the objectivity of human reasoning, a scientific approach and precise measures are used to establish generalizable conclusions (Gratton, 2010). Data collected are often quantitative, so statistical methods can be used (David and Sutton, 2011). Although the objectivity of the positivist approach allows for precise conclusions and statistically-based results, these findings may lack real-world applicability, limited by emotion, context and other factors that influence the human experiences (Scotland, 2012).

In contrast with positivism, interpretivism has its foundations in the ontological position of relativism and the epistemological approach of subjectivism (Scotland, 2012). Interpretivists believe that reality is subjective and influenced by our senses, personal experience, and perspective (David and Sutton, 2011). Absolute knowledge does not exist to be discovered, but rather knowledge can only exist through our interaction with it (Scotland, 2012). Thus, interpretivist research often takes the form of case-studies or qualitative interviews, considering the emotions, experiences and viewpoints of the subjects (Gratton, 2010). Interpretivists would argue that sport is a social science and cannot be understood without taking into consideration the subjective nature of the human experience (Pisk, 2014). This approach allows for the evaluation of more abstract concepts using qualitative methods, but conclusions can be less precise, highly context specific, and limited in generalizability (Scotland, 2012).

Whereas other areas of study may fit more naturally into positivism or interpretivism, sports science can be considered using either or a mixed-methods approach (Pisk, 2014). As sport at its simplest can be considered the physical motion of humans, sports science by nature consists of both a discrete physical motion element (objective) and human component (subjective) (Pisk, 2014). The aim of this research was primarily focused on the physical motion element of sport. As such, it was decided that given the physiological focus and the researcher's objectivist viewpoint that a positivist approach would be most appropriate. Given the primary variables of interest were training load and recovery monitoring measures, summarized using numerical scores, a positivist approach was best suited to the specific measurement and quantitative nature of the research design. Additionally, this approach allowed for statistical analysis and precise summaries of findings. However, despite the overall positivist approach, some consideration was given to the subjective nature of sport. Specifically, although still evaluated using numerical scores, the subjective measures of ratings of perceived exertion and the Recovery Stress-Questionnaire for Athletes were used in addition to objective metrics (Foster *et al.*, 2001; Kellmann and Kallus, 2001). Incorporating these subjective measures allowed for consideration of individuals' experiences, overcoming some of the limitations of a fully positivist approach. Overall, a positivist approach was taken, using quantitative data to make objective measurements, with some consideration given to the subjective nature of athlete load and recovery.

2.2 Research design

2.2.1 Applied sport science research

The goal of sport science research is to further the knowledge base on sport using the scientific process, with the overall aim of improving performance (Bishop, 2008). Since its onset this has resulted in significant changes to sporting practice, specifically at elite levels, and an expanding practical and theoretical understanding of sport (Bishop, 2008; Drust and Green, 2013). For example, multidisciplinary teams are now commonplace in professional sports environments with sport-science related staff filling specialist roles (Drust and Green, 2013; Bartlett and Drust, 2021). In well-funded sports such as football, the scientific approach impacts all aspects of modern organizations, including pre-game, within-game, and post-game strategies, and elements of travel, player development pathways, and overall club structures (Drust and Green, 2013). However, despite this integration, the knowledge transfer in sport science from research into practical settings is still considered to be poor, and the academic community has come under criticism for studying irrelevant problems or promoting practically ineffective solutions (Bishop, 2008; Bartlett and Drust, 2021). In response to this, several frameworks have been developed for improving knowledge translation in sports science (Bishop, 2008; Bartlett and Drust, 2021). Additionally, with increasing research, it has been suggested that there is the need to 'raise the bar' and elevate the quality of the sport specific research that is being performed (Atkinson and Nevill, 2001; Abt *et al.*, 2022).

The research undertaken in this thesis addresses these concerns, with a focus on practical applications to allow for effective knowledge translation. The evidence-based model developed was designed to facilitate ease of transfer from research to applied settings, to improve the practical monitoring of athletes using metrics supported by the literature. Considerations were also given to practical application throughout research design, particularly when developing a new pitch-based protocol for individualized training impulse testing (Chapter 7). The chapters also include 'practical applications' sections, clearly outlining how the findings from the individual studies can be applied in practice. These actions align with the current recommendations for applied research in sport science to increase knowledge transition (Bishop, 2008; Drust and Green, 2013; Bartlett and Drust, 2021). Although not all recommendations for 'raising the bar' have been achieved in this research programme, the rationale for the uniqueness of hockey, the statistical approaches taken, and the focus on diversity via the studying of female hockey athletes, follow the suggested parameters for increasing research quality (Abt *et al.*, 2022).

2.2.2 Methods

Although detailed methods are presented separately for each study, an overview of the research design and approach is outlined here. In order to evaluate athlete monitoring metrics in competitive sport without impacting the athlete experience, a natural experiment approach was taken, with several discrete studies performed. The goal was not to directly influence the training load or recovery status of athletes, but rather to evaluate the typical load of the athletes and the interactions between these metrics and other performance measures. As a result, a controlled experimental design was not appropriate or feasible. Although some may argue that this design limits the external validity of findings, this approach also allowed for the inclusion of more athletes and in their usual in-season environment than could have been achieved if an experimental approach was taken. Furthermore, this design prevented any ethical concerns associated with putting athletes in a deliberate state of overtraining or undertraining, or otherwise predetermining their recovery approaching competition. Due to the nature of athlete monitoring, a repeated measures design was used whereby the same athletes were monitored over a period of time in each individual study. Although repeated measure data were not independent and as such needed to be analyzed appropriately (Bland and Altman, 1995), this allowed for increased statistical power given that squad sizes limited the sample size of the studies.

Rather than one extended study, several discrete studies were performed as part of this research program. The cause of this was two-fold. Firstly, discrete studies allowed for the manipulation and evaluation of monitoring metrics in isolation, minimizing potential confounding variables. Specifically, recovery monitoring, external training load, and internal training load were all specifically evaluated in distinct studies, with some metrics still measured across studies. Secondly, due to the occurrence of COVID-19 during the research period, this allowed for blocks of research to take place in a constantly changing sporting and restrictive landscape. As noted in chapter 3 and in Appendix A, COVID-19 significantly impacted the research undertaken, particularly in terms of the athlete recovery monitoring research.

2.2.3 Participants

Participants in the studies were university hockey athletes competing at the national league level for Durham University. These teams were selected partially due to convenience and also to provide data on an infrequently monitored group of athletes. Where research has been performed on both American university hockey athletes (Astorino *et al.*, 2004; Haydt, Pheasant and Lawrence, 2012; Vescovi and Frayne, 2015; Sell and Ledesma, 2016; Walker *et al.*, 2020) and England hockey league athletes (Stagno, Thatcher and Van Someren, 2007; Macutkiewicz and Sunderland, 2011; Sunderland and Edwards, 2017; Vinson, Gerrett and James, 2018; Noblett *et al.*, 2023), little research has been conducted on British university athletes who participate in both university and adult leagues. The increased demands associated with playing in two leagues highlight the need for appropriate athlete monitoring in this population. Samples changed across the studies given both the nature of the research (some including male and female athletes and others only female) and the makeup of the team over the years. Written informed consent was obtained from all participants, and the Durham University Department of Sport and Exercise Science Ethics Committee approved all protocols and safety procedures for the research.

2.2.4 Statistical Approach

A variety of statistical techniques were used given the nature of the research questions and data collected. A primarily frequentist approach was taken to statistical analysis, with p-values reported. However, the limitation of traditional null hypothesis significance testing p-values was considered, with these values frequently misinterpreted (Greenland *et al.*, 2016; Wasserstein, Schirm and Lazar, 2019; Lakens *et al.*, 2020). Specifically, with substantial sample sizes, as are often the case with repeated measures data, differences may be statistically significant despite being so small as to have no meaningful, practical application. Therefore, the smallest meaningful change was considered, and p-values were calculated via the minimum effects test to ensure findings were not only statistically but also practically significant (Lakens, 2017; Wasserstein, Schirm and Lazar, 2019; Lakens et al., 2020). The range of no practical significance was dependent on the hypothesis being tested. For example, when evaluating correlations across training load measures where an association was to be expected, the minimum effects tests was used to test for a strong correlation, with p-values then having a more practically meaningful interpretation. In all cases, data were first checked for normality using visual inspection of Q-Q-plots, stem-and-leaf charts and histograms. Given the robustness of the statistical techniques used, the limitations of p-values, and the quantity of repeated measures data, these metrics were chosen in favor of the null hypothesis tests of Shapiro-Wilks or Kolmogorov-Smirnov. Results were presented as means and standard deviations, and statistical significance was set at $p < 0.05$ to favor minimizing type II errors and allow for the exploration of all potential associations. Full details of all procedures and analyses are provided within each empirical chapter.

Chapter 3: Summary of Literature – Athlete Recovery Monitoring

3.1 Introduction to the training process

Sport training is the process of performing physical activity to elicit an improvement in physical or sport-specific abilities (Coutts, Crowcroft and Kempton, 2017; Impellizzeri, Marcora and Coutts, 2022). The exercise performed results in a physiological response which provides the stimulus for potential positive adaption (Impellizzeri, Marcora and Coutts, 2022). To be effective, training must be specific to the nature of the intended outcome (Impellizzeri, Rampinini and Marcora, 2005; Bangsbo, 2015). Specifically, the quality, quantity, and organization of the training performed will impact the physiological response and thus the stimuli for changes in physical abilities (Impellizzeri, Rampinini and Marcora, 2005). For example, endurance and speed training are different stimuli eliciting different physiological responses and thereby resulting in distinct adaptations (Bangsbo, 2015). Additionally, although training will result in an acute physiological response, for chronic adaptation, training must be continued for an extended period of time (Banister, 1991; Coutts, Crowcroft and Kempton, 2017).

An athlete's training can be considered both in terms of the amount of physical work performed and the physiological response caused by that work, also known as external and internal training load. External training load is the amount of physical work performed regardless of physiological response and can be thought of as the quality, quantity, and organization of a training session. External training load is commonly measured via global positioning systems or accelerometers and includes metrics such as total distance covered, average speed, distance in speed zones, and accelerations/decelerations (Bourdon, 2017). Measuring a specific physical output, external training load metrics are clearly defined, and their validity can be assessed through comparison with criterion values (Weaving *et al.*, 2017). On the other hand, internal training load measures aim to summarize the body's physiological response to exercise, based on a specific response such as perceived exertion or heart rate (Fox *et al.*, 2018). However, although protocols for the measurement of these physiological markers have been clearly established, there is a lack of consensus on how to best summarize these data into training load scores (Weaving *et al.*, 2017; Gonzalez-Fimbres *et al.*, 2019; Fox *et al.*, 2018; McLaren *et al.*, 2018).

When an athlete completes a training session or bout of exercise, they will perform a certain amount of physical work (external training load), characterized by organization, periodization, quality and quantity of the work (Coutts, Crowcroft and Kempton, 2017; Impellizzeri, Marcora and Coutts, 2022). Depending on individual and contextual factors, the external load performed will elicit a physiological response (internal training load) distinct to the individual (Jeffries *et al.*, 2020; Impellizzeri, Marcora and Coutts, 2022). The training status of the individual will determine the physiological response to a physical load. For example, running an eight-minute mile (external load) will elicit a different physiological response (internal load) on an untrained individual compared to a university athletics athlete. Therefore, it is the internal load rather than the external load that will determine the training effects and performance outcomes (Impellizzeri, Rampinini and Marcora, 2005; Coutts, Crowcroft and Kempton, 2017; Gabbett *et al.*, 2017; Jeffries *et al.*, 2020; Impellizzeri, Marcora and Coutts, 2022). Additionally, outside individual factors such as the recovery, nutrition, psychological status, and genetics, will impact any adaption that occurs (Impellizzeri, Marcora and Coutts, 2022).

Internal training load has both acute and chronic, negative and positive, training effects (Jeffries *et al.*, 2020). Specifically, the impact of training on the body can be considered in terms of Banister's fitness-fatigue model, illustrating the relationship between dose and recovery (Banister, 1991; Coutts, Crowcroft and Kempton, 2017). The stimulus from internal load will impact both an athletes' fitness and fatigue levels, with predicted performance expressed as the mathematical difference of fitness and fatigue (Banister, 1991). Given the balance of stress and recovery over time, these acute positive and negative impacts on fitness and fatigue can result in chronic positive or negative training effects. As fatigue decays faster than fitness, when periods of intense training are followed with adequate rest, performance will improve beyond baseline levels, known as the supercompensation effect (Coutts *et al.*, 2007; Coutts, Crowcroft and Kempton, 2017). However, if there is insufficient rest and recovery, the impact of fatigue can outweigh positive impacts on fitness, resulting in decremental training outcomes associated with overtraining and under recovery (Kellmann, 2010; Coutts, Crowcroft and Kempton, 2017). This chapter will focus on this element of the training process, specifically considering athlete recovery and how monitoring athlete recovery can be performed to help minimize negative training outcomes.

3.2 Athlete recovery

3.2.1 Introduction to athlete recovery

The vast majority of published data on athlete monitoring has focused on measuring performance and load in a training or competition setting (Saw, Main and Gastin, 2016; Taylor *et al.*, 2017; Fox *et al.*, 2018). However, even amongst elite performers, athletes only spend a relatively small percentage of their total time completing sports-specific activities (Sperlich and Holmberg, 2017). Consequently, it is impossible to fully determine how athletes are progressing and responding to training if time spent outside of their sporting environment is not considered. The choices that athletes make throughout the day, such as decisions regarding nutrition, sleep, and activity levels, as well as other roles, responsibilities and stressors in their lives, all impact athletes' physiological and physical responses to training (Sperlich and Holmberg, 2017). Therefore, if monitoring solely focuses on measuring variables such as training load, which only provide information on what occurs in an on-pitch environment, coaches and sports-scientists will not be able to accurately determine the recovery status and wellbeing of their athletes (Nässi *et al.*, 2017b; Kraft *et al.*, 2018). Consequently, monitoring athletes' off-pitch recovery and overall wellbeing is critical for accurately determining individualized training dose prescriptions (Lambert and Borresen, 2006; Nässi *et al.*, 2017b; Kraft *et al.*, 2018).

Before recovery monitoring can be discussed, it is important to first define recovery. Kellmann and Kallus, leading researchers in the area of athlete fatigue and recovery, define recovery as "an inter-individual and intra-individual multi-level (eg. psychological, physiological, social) process in time for the re-establishment of performance abilities" (Kellmann and Kallus, 2001). As demonstrated by this definition, recovery is a complex and multifaceted process, not simply a set of activities or exercises performed by an athlete (Kellmann and Kallus, 2001). Recovery has many different elements, including both psychological and physiological mechanisms, with the unifying aspect of all components being a contribution towards a return in performance levels (Kellmann and Kallus, 2001). Stress and recovery work in contrast, with recovery consisting of a reduction, change or interruption of stress, and, as a result, recovery is highly situational and individualized (Kellmann, 2010). Recovery has been classed as an umbrella term covering various modalities such as physiological regeneration and psychological rest (Duffield *et al.*, 2018). Regeneration refers to the physiological component of recovery and encompasses activities such as sleep and massage, whereas the psychological aspects of recovery focus on rest from the mental fatigue of sport and includes techniques such as relaxation exercises (Duffield *et al.*, 2018). Recovery can also be classed as passive, active, and proactive with active methods, such as stretching, aimed at responding to the metabolic impacts of fatigue, passive methods involving a state of inactivity and rest, and proactive methods consisting of the choices made by athletes outside of sport, such as decisions regarding alcohol consumption (Duffield *et al.*, 2018). Overall, recovery refers to a variety of different activities, choices, and actions performed by athletes outside of their direct participation in sport that impact athletic performance (Kellmann and Kallus, 2001).

3.2.2 Importance of recovery monitoring

In order to achieve maximal benefits from training and to perform at one's best, a balance must be maintained between training stress and recovery (Kellmann, 2010). Although successful training requires overload, this overload is only beneficial when accompanied with adequate recovery (Lambert and Borresen, 2006; Meeusen *et al.*, 2013). Over time, if training loads are constantly elevated and athletes are given insufficient time to recover, an imbalance will occur between stress and recovery, causing extreme fatigue, overreaching, eventually overtraining (Kellmann, 2010). Athletes in these maladaptive states will experience performance declines and negative mood changes, for which there is no treatment apart from adequate rest (Meeusen *et al.*, 2013). As athletes respond differently to identical training programs and choices made outside of training time impact recovery, recovery status should be considered for each athlete individually (Kölling *et al.*, 2015). The same training load that results in positive adaptation in one athlete may cause overtraining in another, so even the most carefully planned training program will not be best suited to all members of a team (Lambert and Borresen, 2006; Kölling *et al.*, 2015). Furthermore, if individualized athlete recovery and wellbeing are not monitored, it is often impossible to know when an athlete is experiencing training distress until notable performance decrements have occurred, at which point recovery may take weeks or months (Meeusen *et al.*, 2013). Therefore, regular recovery monitoring is important to ensure that athletes are maintaining an appropriate balance between training and recovery, and instances of under-recovery are identified before overtraining occurs (Lambert and Borresen, 2006; Kellmann, 2010).

Some coaches may claim that recovery monitoring is unnecessary because they can use their experience and expertise to evaluate athlete wellbeing and identify early warning signs of overtraining (Pope, Penney and Smith, 2018). However, as is the case with monitoring training load, where it has been shown that coaches are unable to accurately determine the work performed without training load measures (Bompa, 1999; Brink *et al.*, 2014), it has also been shown that coaches are unable to accurately assess athlete recovery states based on their intuition alone (Kraft *et al.*, 2018). Specifically, when US university head coaches in volleyball, football, and basketball were asked to evaluate individual athlete recovery following warmup in 433 training sessions, coaches significantly overestimated recovery, compared to athlete's self-reported scores ($p < 0.05$) (Kraft *et al.*, 2018). Furthermore, trainings sessions that coaches designed to be 'light' or 'moderate,' were perceived to be harder than intended ($p < 0.05$), and 'hard' sessions were perceived to be lighter than intended ($p < 0.05$), demonstrating a trend toward monotony, which further contributes to overtraining (Kraft *et al.*, 2018). A study interviewing elite rowing coaches found that the cues and intuition that the coaches used to evaluate fatigue and recovery had little relation to overtraining markers validated in the literature (Pope, Penney and Smith, 2018). Additionally, the study found that when specific recovery protocols and questionnaires were not put in place, coaches are often hesitant to ask athletes about their life outside of sport so as not to appear overbearing (Pope, Penney and Smith, 2018). Therefore, coaches often rely solely on external cues when determining athletes' recovery status (Lambert and Borresen, 2006; Kraft *et al.*, 2018). However, as fatigue and overtraining have both physiological and psychological components, even the most experienced coaches will not be able to fully assess athletes' recovery states without receiving athlete input (Kraft *et al.*, 2018).

Recovery monitoring is of heightened importance in vulnerable groups of athletes, such as university athletes and athletes at risk of eating disorders (Meeusen *et al.*, 2013; Hamlin *et al.*, 2019). A study of 182 elite university athletes found that athletes were significantly more susceptible to stress and illness at certain times of the year, due to their academic commitments ($p < 0.05$) (Hamlin *et al.*, 2019). Although the increased stress was academic rather than sport-related, the stress still impacted athletes' physiological and psychological responses to training and placed them at an elevated risk for overtraining and injury (Hamlin *et al.*, 2019; Nicolas *et al.*, 2019). University athletes frequently have inadequate sleep schedules, particularly when academic stresses are high, and even moderate sleep deprivation has been shown to be detrimental to both physiological and psychological recovery after exercise (Claudino *et al.*, 2019; Hamlin *et al.*, 2019). Therefore, it is critical for coaches working with university athletes to be mindful of students' academic demands to ensure that the combination of academic and athletic stressors do not negatively impact performance or wellbeing (Hamlin *et al.*, 2019). Additionally, although the performance declines associated with under-recovery are distinct from those caused by eating disorders, since nutrition is a key element of recovery, athletes who suffer from disordered eating are more susceptible to overtraining and under-recovery (Meeusen *et al.*, 2013). Specifically, the combination of carbohydrate depletion and multiple days of intense training has been shown to result in overtraining symptoms, and dehydration exacerbates the stress response (Eichner, 1995; Meeusen *et al.*, 2013). Therefore, athletes who are at risk for negative energy balance need regular monitoring to ensure that they achieve adequate nutrition, allowing for recovery and a positive training response (Eichner, 1995; Meeusen *et al.*, 2013).

In summary, just as appropriate training is required for positive athletic adaptation, adequate recovery must also be achieved to ensure that athletes maintain a long-term balance between stress and recovery. As athlete recovery cannot be accurately predicted by coaches and varies based on non-sports factors, regular, individualized recovery monitoring is critical to ensure that athletes are not in a persistent state of under-recovery, which can result in overreaching and overtraining.

3.3 Overreaching and overtraining

3.3.1 Defining overreaching and overtraining

The principle of progressive overload indicates that successful training programs must incorporate some element of overload (Meeusen *et al.*, 2013; Duffield *et al.*, 2018). Intense training sessions are regularly used to improve performance by overloading athletes, and these sessions are often followed by acute symptoms of fatigue and mild drops in performance (Meeusen *et al.*, 2013; Duffield *et al.*, 2018). During appropriate periodization, when intense training blocks are followed by sufficient rest and recovery, performance is enhanced beyond baseline levels, as a result of the supercompensation effect (Coutts, Slattery and Wallace, 2007; Kellmann, 2010; Nässi *et al.*, 2017b). This type of overload training is often termed functional overreaching, as short-term performance decline and fatigue occur; however, when followed with sufficient recovery, performance improves (Meeusen *et al.*, 2013; Pope, Penney and Smith, 2018). When excessive training is continued for extended periods of time, without adequate recovery, athletes enter a maladaptive state of either non-functional overreaching or overtraining, in which performance stagnates or declines (Nässi *et al.*, 2017b). In 2013, the European College of Sports Science and the American College of Sports Medicine released a joint consensus statement on overtraining syndrome in which they presented an "Overtraining Continuum" as reproduced in Figure 3.1 (Meeusen *et al.*, 2013). This continuum illustrates the various stages of overreaching and overtraining syndrome, and demonstrates the potential progression of these outcomes as training intensifies and continues without adequate recovery (Meeusen *et al.*, 2013).

PROCESS	TRAINING (overload)	INTENSIFIED → TRAINING		
OUTCOME	ACUTE FATIGUE	FUNCTIONAL OR (short-term OR)	NON-FUNCTIONAL OVERREACHING (extreme OR)	OVERTRAINING SYNDROME (OTS)
RECOVERY	Day(s)	Days – weeks	Weeks – months	Months - …
PERFORMANCE	INCREASE	Temporary performance decrement (e.g., training camp)	STAGNATION DECREASE	DECREASE

Figure 3.1 Overtraining Continuum (Meeusen et al., 2013)

As demonstrated by the continuum, the relationship between the stages of functional overreaching, non-functional overreaching, and overtraining syndrome are fluid and difficult to distinguish (Meeusen *et al.*, 2013). On one end of the spectrum, functional overreaching leads to performance improvements via supercompensation after only short-term performance decrements without lasting negative effects (Coutts, Slattery and Wallace, 2007; Meeusen *et al.*, 2013). As a result, periods of functional overreaching are often deliberated incorporated into programs as a regular component of training (Coutts, Slattery and Wallace, 2007). However, in some cases, when athletes experience the short-term performance decrements caused by functional overreaching, they assume the decline is caused by a lack of fitness or skill and respond by attempting to train harder, rather than allowing for adequate recovery (Tobar, 2005). This pushes athletes farther along the continuum towards non-functional overreaching and overtraining syndrome, depending on the length and severity of the stress and recovery imbalance (Nässi *et al.*, 2017b). At this point, non-functional overreaching and overtraining syndrome can often only be differentiated retrospectively, as they are distinguished not by the symptoms, but rather by the length of the recovery time, with non-functional overreaching requiring weeks or months and overtraining syndrome requiring months or years (Coutts, Slattery and Wallace, 2007; Meeusen *et al.*, 2013).

3.3.2 Impacts of overreaching and overtraining

The symptoms of non-functional overreaching and overtraining syndrome are multifaceted, with both physiological and psychological changes occurring (Tobar, 2005; Coutts, Slattery and Wallace, 2007; Kellmann, 2010; Meeusen *et al.*, 2013; Nässi *et al.*, 2017b). The most-prominent symptom of non-functional overreaching and overtraining is a sport-specific drop in performance that continues despite an extended period of recovery (Meeusen *et al.*, 2013). Psychologically, this performance decrease is paired with reduced wellbeing and a disturbance in mood state, often manifested through general apathy, decreased self-esteem, irritability, or boredom (Kellmann, 2010; Meeusen *et al.*, 2013; Nässi *et al.*, 2017b). Athletes suffering from non-functional overreaching and overtraining syndrome report symptoms similar to depression and, in some instances, are clinically depressed (Tobar, 2005; Saw, Main and Gastin, 2016). Physiologically, in addition to decreased performance, the symptoms of non-functional overreaching and overtraining syndrome include depressed immune function, hormonal dysregulation, and sleep disturbances (Coutts *et al.*, 2007; Kellmann, 2010; Nässi *et al.*, 2017b). Furthermore, these symptoms may be accompanied by a lack of appetite and weight loss, which can further hinder athlete recovery (Kellmann, 2010). Notably, as there is no one tool to definitely identify overtraining syndrome and non-functional overreaching, diagnoses are often made by ruling out other potential causes of an athlete’s symptoms (Eichner, 1995; Hagstrom and Shorter, 2018; Pope, Penney and Smith, 2018). For example, endocrinological disorders, infectious diseases, iron deficiency, and eating disorders should all be excluded as possible causes of performance decrements and mood disturbances (Meeusen *et al.*, 2013). Once non-functional overreaching or overtraining syndrome is diagnosed, the only treatment is rest and recovery, the timeframe for which is highly individualized and very difficult to predict (Tobar, 2005; Meeusen *et al.*, 2013).

Although the many symptoms and long timeframe for recovery from overtraining syndrome may suggest that this condition is extreme and uncommon, the prevalence of overtraining has been found to be high, especially among elite athletes (Morgan *et al.*, 1987b; Morgan, 1988; Kentta, Hassmen and Raglin, 2001; Gould *et al.*, 2002; Meeusen *et al.*, 2013). A survey of distance runners found that 60% of female and 64% of male elite runners experienced at least one case of non-functional overreaching or overtraining in their career, with non-elite runners reporting a career rate of 33% (Morgan *et al.*, 1987b; Morgan, 1988). Additionally, across 296 team and individual-sport athletes from the 1996 Atlanta Olympics, the self-reported rate of overtraining was 23% (Gould *et al.*, 2002). This finding may appear contradictory to the data on elite runners; however, it has been shown that non-functional overreaching and overtraining are significantly more common in individual-sport athletes than in team-sport athletes (Kentta, Hassmen and Raglin, 2001). Specifically, a study of 272 elite age-group athletes found that non-functional overreaching and overtraining were more common in individual-sport athletes (48%) than team-sport athletes (30%) (Kentta, Hassmen and Raglin, 2001). From these studies, there is clearly a lack of consensus on the prevalence of non-functional overreaching and overtraining across various athlete populations; however, this is likely due to the difficulty of using retrospective surveys to determine prevalence, interpretations of overtraining varying, and athletes self-reporting rather than being diagnosed by a professional (Meeusen *et al.*, 2013). Nevertheless, regardless of the study considered, the high prevalence of non-functional overreaching and overtraining and the multifold negative symptoms associated with these conditions indicate the importance of prevention. As such, any coach working with high performance athletes, particularly those aiming to maximize performance through periodization and training load monitoring would be unwise not to consider the potential occurance of non-functional overreaching and overtraining in their athletes.

As the only treatment for non-functional overreaching and overtraining is rest and recovery, the goal of many coaches and medical professionals is the prevention and early recognition of the potential warning signs of overtraining (Eichner, 1995; Tobar, 2005; Meeusen *et al.*, 2013). If the signs and symptoms of overreaching are identified as they start to appear, athletes can be prescribed the appropriate rest and recovery to prevent them declining further into overtraining syndrome (Coutts, Slattery and Wallace, 2007; Meeusen *et al.*, 2013). Despite widespread research in the area, no one tool has been identified as a definitive marker for athlete recovery status and a diagnostic indicator of impending non-functional overreaching or overtraining (Eichner, 1995; Tobar, 2005; Hagstrom and Shorter, 2018; Pope, Penney and Smith, 2018). However, a variety of promising tools have been developed and markers have been identified to give early warning signs for a variety of overreaching and overtraining symptoms and together these tools can provide information on how athletes are responding to a given training dose. (Morgan *et al.*, 1987a; Rushall, 1990; Freitas *et al.*, 2014; Bellenger *et al.*, 2016a; Kallus and Kellmann, 2016; Nässi *et al.*, 2017a). These monitoring methods can be broken down into two broad categories, subjective, perceptual measures, consisting of questionnaires and direct feedback from athletes, and objective measures, such as blood-based markers and measures of neuromuscular fatigue and autonomic nervous system function (Morgan *et al.*, 1987a; Rushall, 1990; Freitas *et al.*, 2014; Bellenger *et al.*, 2016a; Kallus and Kellmann, 2016; Nässi *et al.*, 2017a). When athletes experience imbalance between training or life stresses and recovery, the combination of performance stagnation or decline and negative results on subjective and/or perceptual recovery measures provides evidence that overreaching or overtraining may have occurred and steps should be taken to ensure that appropriate recovery is achieved (Eichner, 1995; Tobar, 2005; Meeusen *et al.*, 2013). Some of these tools, particularly heart rate-based measures have also been used to monitor athlete training status in terms of fitness and preparedness to perform (Buchheit, 2014; Daanen *et al.*, 2012; Schneider *et al.*, 2018). However, the use of these measures for fitness assessments is distinct from monitoring athletes' stress and recovery levels, and, as such, will not be addressed in this review of recovery monitoring.

3.4 Subjective recovery measures

Subjective recovery monitoring is based on self-reported perceptual measures of recovery status and overall wellbeing. Perceptual measures can incorporate mood status, recovery modalities, behavioral and physical symptoms, overall health, and mental wellbeing; however, the majority of subjective recovery measures are questionnaires focused on the psychological status of athletes (Saw, Main and Gastin, 2016). Overtraining and overreaching result in emotional disturbances, impacting athletes' mood and psychological wellbeing; therefore, many subjective measures of athlete recovery focus on monitoring emotional changes and psychological disruptions that indicate early signs of overtraining (Morgan *et al.*, 1987a; Rushall, 1990; Kellmann and Kallus, 2001; Meeusen *et al.*, 2013; Saw, Main and Gastin, 2016; Nässi *et al.*, 2017b). Additionally, as the aim of recovery monitoring is to ensure that athletes maintain an appropriate balance between stress and recovery, some subjective recovery measures directly ask athletes to report their perceived stress and recovery levels (Nässi *et al.*, 2017b). Over time, a variety of questionnaires have been developed for subjectively monitoring athlete recovery status and providing early warning signs for potential overreaching and overtraining (Morgan *et al.*, 1987a; Rushall, 1990; Kellmann and Kallus, 2001; Meeusen *et al.*, 2013; Saw, Main and Gastin, 2016; Nässi *et al.*, 2017b).

There are many benefits to using subjective measures to monitor athlete recovery. Firstly, questionnaires are inexpensive and easy to administer (Saw, Main and Gastin, 2016). Compared to objective measures, which often require specialized laboratory equipment and standardized conditions, the insignificant cost and lack of equipment needed make subjective measures very practical for use in applied settings (Robson-Ansley, Gleeson and Ansley, 2009; Meeusen *et al.*, 2013; Archbold *et al.*, 2018). Additionally, unlike many tests of physical performance, subjective questionnaires are not physiologically taxing, ensuring that the athlete monitoring itself does not further contribute to physiological stress and under-recovery (Meeusen *et al.*, 2013). Subjective measures are also particularly useful in team-sport settings, as questionnaires can be provided to a large group of athletes at once and results can be obtained almost instantaneously (Nässi *et al.*, 2017b; Nicolas *et al.*, 2019). Furthermore, asking athletes about their recovery has been shown to increase athlete awareness of the sources and symptoms of stress, resulting in improved recovery activities (Robson-Ansley, Gleeson and Ansley, 2009). In addition to the practical benefits, the research also suggests that subjective measures are better suited than objective measures for measuring the earliest indications of overtraining and under-recovery (Saw, Main and Gastin, 2016; Nässi *et al.*, 2017b). Specifically, a systematic review of 56 studies reported that subjective recovery measures responded to overreaching with increased sensitivity and superior consistency than objective measures (Saw, Main and Gastin, 2016). Psychological reactions occur with small changes in training load and recovery, with psychological symptoms occurring before biochemical and hormonal changes appear (Nässi *et al.*, 2017b). Therefore, as psychological symptoms are one of the earliest signs of overreaching, monitoring psychological changes is one of the best methods of proactively identifying overreaching and under-recovery (Lambert and Borresen, 2006).

Despite the many advantages, there are also several key drawbacks of subjective recovery monitoring. As subjective monitoring is based on self-reported responses, the results of these questionnaires can be impacted by response distortion, with athletes intentionally misreporting or faking their responses (Meeusen *et al.*, 2013; Kölling *et al.*, 2015). For example, athletes may exaggerate stress and understate their recovery in an attempt to have coaches reduce the intensity of future training sessions, or report falsely high levels of recovery in an effort to impress their coaches (Kölling *et al.*, 2015). Therefore, care should be taken to ensure that subjective measures are appropriately and consistently administered, and athletes should be informed that results will not be used for selection purposes in order to decrease the risk of response bias (Meeusen *et al.*, 2013; Kölling *et al.*, 2015). Furthermore, the frequency of administration and time required for the completion of questionnaires should be managed in order to minimize the burden on athletes and maximize compliance (Gabbett *et al.*, 2017). It is beyond the scope of this review to examine all subjective recovery measures mentioned in the literature, so the most notable and well-studied questionnaires will be discussed below (Robson-Ansley, Gleeson and Ansley, 2009).

3.4.1 Profile of Mood Scores (POMS)

Developed in 1971, the first questionnaire regularly used to monitor psychological symptoms in athletes was the Profile of Mood Scores (POMS) (Morgan *et al.*, 1987a). POMS consists of 65 items reported on a 5-point Likert-scale, with questions corresponding to distinct subscales for tension, depression, anger, fatigue, confusion, and vigor (Morgan *et al.*, 1987a). Overall mood status scores are computed by summing the five negative mood states, adding 100, and subtracting the positive state (vigor) (Morgan *et al.*, 1987a). POMS may be used with a variety of time sets, including ‘in the past week,’ ‘today,’ and ‘right now’ (Leunes and Burger, 2000). Although the questionnaire was initially designed for use in counselling, its function has evolved, and it is now regularly used for monitoring athletes (Lambert and Borresen, 2006). POMS has been shown to have acceptable predictive and concurrent validity, and internal consistency of the six subscales ranged from $\alpha = 0.84 - 0.95$ in a study of 2000 psychometric outpatients (Leunes and Burger, 2000). Additionally, the results of POMS have been shown to be consistent across sexes with no significant differences ($p > 0.05$) occurring between male and female athletes except when training programs differed (Morgan *et al.*, 1987a).

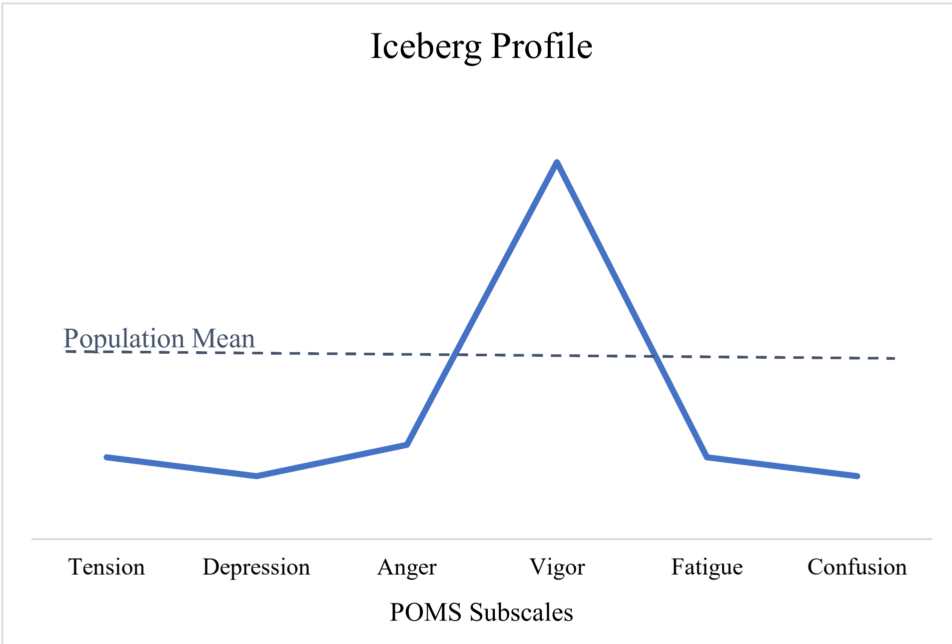


Figure 3.2 Profile of Mood Scores Iceberg Profile

POMS has been studied in large samples of athletes across various sports, and changes in mood state have been repeatedly shown to be sensitive to changes in training load and an early indicator of overreaching and overtraining (Gutmann *et al.*, 1984; Morgan *et al.*, 1987a; Morgan *et al.*, 1987b; Raglin, Morgan and Luchsinger, 1990; Verde, Thomas and Shephard, 1992; Berglund and Safstrom, 1994; O'Connor, Raglin and Morgan, 1996; Filaire *et al.*, 2001; Uusitalo, 2001; Halson, 2014). In general, athletes score above average on vigor and below average on the five negative mood states, in what has been termed an 'iceberg profile' (Figure 3.2) (Morgan *et al.*, 1987a; Eichner, 1995). However, in response to increased training stress, scores for vigor have been shown to decrease with concurrent increases in negative mood states in swimming (Morgan *et al.*, 1987a; Morgan *et al.*, 1988), running, (Verde, Thomas and Shephard, 1992; Raglin and Morgan, 1994), speedskating (Gutmann *et al.*, 1984) rowing (Raglin, Morgan and Luchsinger, 1990), canoeing (Berglund and Safstrom, 1994), soccer (Filaire *et al.*, 2001), and triathlon (O'Connor, Raglin, and Morgan, 1996). When the training stimulus is decreased, overall mood status increases, with improved mood states measured on POMS paralleling improved performances in athletes recovering from overtraining syndrome (Morgan *et al.*, 1987a; Robson-Ansley, Gleeson and Ansley, 2009). Additionally, in a four-week study of elite swimmers with a double-blind procedure, the results of POMS were used to accurately predict training maladaptation 81% of the time (Morgan *et al.*, 1988).

Although one of the best studied subjective methods of recovery monitoring, review articles have demonstrated that the majority of research on POMS has been performed on individual rather than team-sport athletes (Leunes and Burger, 2000; Nässi *et al.*, 2017b). Additionally, POMS has received criticism for not being sport-specific and for its length, making it impractical for frequent use in non-research settings (Robson-Ansley, Gleeson and Ansley, 2009; Nässi *et al.*, 2017b). With only one positive mood state, POMS overemphasizes the negative, and it does not provide information on the full range of positive emotions associated with recovery (Nässi *et al.*, 2017b; Lundqvist and Kenttä, 2010). As a result, the Emotional Recovery Questionnaire (EmRecQ) has been developed as a 22-item supplement to POMS, focusing on positive emotions to provide a more complete profile of athletes' moods (Lundqvist and Kenttä, 2010). However, adding this supplement only lengthens the already long POMS questionnaire, and only one study, based on a single ultrarunner, has utilized this questionnaire (Johnson *et al.*, 2016). Therefore, more research is needed to validate and justify the use of this supplement.

3.4.2 Daily Analyses of Life Demands for Athletes (DALDA)

In 1990, the Daily Analyses of Life Demands for Athletes (DALDA) questionnaire was developed as a sports-specific method of measuring athletes' sources and symptoms of stress (Rushall, 1990). DALDA consists of two parts, part A, which is made up of 9 questions relating to sources of stress, and part B, which considers 25 symptoms of stress (Rushall, 1990). Symptoms and stressors are scored on a 3-point scale of 'worse than normal,' normal,' and 'better than normal,' and results are commonly expressed graphically to allow for ease of comparison over time (Rushall, 1990; Robson-Ansley, Gleeson and Ansley, 2009). As implied in the name, DALDA was designed to be completed daily; however, follow-up studies have indicated that weekly completion does not reduce the sensitivity of the questionnaire to variations in training load (Robson-Ansley, Blannin and Gleeson, 2007). Shorter than POMS and specific to athletes, DALDA was designed to provide a new questionnaire for the detection of early signs of overreaching and overtraining (Rushall, 1990; Robson-Ansley, Gleeson and Ansley, 2009).

DALDA has been shown to be sensitive to changes in the physiological and psychological symptoms of stress that occur as a result of variations in training load (Rushall, 1990; Halson *et al.*, 2002; Coutts, Slattery and Wallace, 2007; Robson-Ansley, Blannin and Gleeson, 2007). In response to increased training stressors and deliberate overreaching protocols, athletes have been shown to report more 'worse than normal' responses for symptoms of stress (Rushall, 1990; Halson *et al.*, 2002; Coutts, Slattery and Wallace, 2007; Robson-Ansley, Blannin and Gleeson, 2007; Milanez *et al.*, 2014), and, in triathletes, a significant increase in negative responses ($p < 0.01$) was shown to occur simultaneously to a significant decrease in immune function ($p < 0.01$) (Robson-Ansley, Blannin and Gleeson, 2007). In terms of performance, a systematic review found strong evidence of a negative association between sources of stress and sustained performance (Saw, Main and Gastin, 2016). Therefore, it has been recommended that if an athlete reports an increased number of 'worse than normal' responses on three consecutive days that action be taken to address the potential lack of balance between stress and recovery (Rushall, 1990; Robson-Ansley, Gleeson and Ansley, 2009; Nässi *et al.*, 2017b). With sources of stress reported in part A of the questionnaire, it is easy to determine whether sport-related or other life stressors may be causing the increase in stress-related symptoms (Rushall, 1990; Robson-Ansley, Gleeson and Ansley, 2009; Nässi *et al.*, 2017b). Furthermore, studies on cyclists and triathletes have shown that DALDA can be used to accurately distinguish, based on significant increases in 'worse than normal' responses in part B ($p < 0.05$), between those athletes positively adapting to training and those reaching states of non-functional overreaching (Halson *et al.*, 2002; Coutts, Slattery and Wallace, 2007). DALDA can also be used to track athlete recovery from overreaching and overtraining, with athletes' scores significantly improving ($p < 0.05$) following a decrease in training after an overload period (Halson *et al.*, 2002; Coutts, Slattery and Wallace, 2007; Robson-Ansley, Blannin and Gleeson, 2007).

One of the top three questionnaires used in the recovery monitoring literature (Saw, Main and Gastin, 2016), DALDA has several key strengths, but also notable weaknesses. Firstly, DALDA is sport-specific and provides information not only on symptoms of stress but also on the causes of stress (Rushall, 1990). Additionally, baseline values do not need to be established, as a comparison with an individual's 'normal' is inherent in the responses (Saw, Main and

Gastin, 2016). From a practical perspective, eliminating the need for baseline measurement saves both time and effort, allowing for quicker implementation (Saw, Main and Gastin, 2016). The DALDA questionnaire is relatively short, with only three possible answers to each question, allowing for fast completion and improving athlete compliance (Robson-Ansley, Gleeson and Ansley, 2009; Nässi *et al.*, 2017b). However, despite this fact, likely due to the daily nature of the questionnaire, no study has tracked DALDA responses in athletes over an extended period of time (Robson-Ansley, Gleeson and Ansley, 2009; Saw, Main and Gastin, 2016). Finally, while elite swimmers were studied to establish the reliability and validity of DALDA, using a test-retest study during the development process of the questionnaire, the statistical outcomes were not reported (Rushall, 1990).

3.4.3 Recovery-Stress Questionnaire for Athletes (RESTQ-S)

The Recovery-Stress Questionnaire (RESTQ) is a detailed questionnaire designed to assess the psychophysical state of an individual by measuring the prevalence of stress and recovery related feelings and actions over a set period of time (Kallus and Kellmann, 2016). It has been adapted to include specific versions for athletes, coaches, children/adolescents, and working individuals (Kallus and Kellmann, 2016). The Recovery-Stress Questionnaire for Athletes (RESTQ-S) is multidimensional, incorporating information on the physical, behavioral, social, and subjective aspects of both stress and recovery, and the structure of RESTQ-S is modular, with 76 questions making up 19 scales - 10 for stress and 9 for recovery (Kallus and Kellmann, 2016). Twelve of the scales are generic, used across all RESTQ questionnaires, with 7 scales specific to sport (Kallus and Kellmann, 2016). The questions on RESTQ-S were designed to assess athletes' stress level and current capacity for recovery, and answers are recorded on a 7-point Likert scale from never (0) to always (6) (Davis, Orzeck and Keelan, 2007). The timeframe of reference for the assessment is typically 3-days; however, 7 and 14-day recalls have also been used in some settings (Nässi *et al.*, 2017b). Comparisons with POMS, the State-Trait-Anxiety Inventory, and the Multidimensional Physical Symptoms List were performed with results indicating acceptable discriminant and convergent validity (Nässi *et al.*, 2017b), and several studies have shown strong positive correlations between RESTQ-S and POMS scales (Kellmann and Klaus-Dietrich, 2000; Davis, Orzeck and Keelan, 2007). However, when psychometric item evaluations were performed on the individual questions, the 2-factor structure and 19 subscales came into question

(Davis, Orzeck and Keelan, 2007). Nevertheless, it was determined that RESTQ-S is still a valid measure of under-recovery, with results supporting the practical applications of RESTQ-S for athlete monitoring (Davis, Orzeck and Keelan, 2007).

RESTQ-S has been shown to be an effective method of measuring athletes' responses to training in a variety of settings across sports (Kellmann and Klaus-Dietrich, 2000; Kellmann *et al.*, 2001; Jurimae *et al.*, 2002; Coutts, Wallace and Slattery, 2007; Nunes *et al.*, 2014; Nässi *et al.*, 2017b). Specifically, a systematic review of various recovery measures found RESTQ-S to be the only subjective measure sensitive to changes in both acute and chronic training load, particularly when the *Fatigue*, *Physical Recovery*, *General Wellbeing*, and *Being in Shape* subscales were considered (Saw, Main and Gastin, 2016). Other subscales were not found to be as responsive; however, when overall stress and recovery scores were determined by combining the subscales, there was sufficient evidence to suggest that scores were still responsive to changes in training load (Saw, Main and Gastin, 2016). In rowers during a period of intense training, several RESTQ-S subscale responses were notably correlated with cortisol levels ($r = 0.65 - 0.76$) (Jurimae *et al.*, 2002), and, in basketball, recovery-stress state was shown to be significantly impaired ($p < 0.05$) in a period of overload training and to return to baseline following a 2-week taper period (Nunes *et al.*, 2014). Most importantly, RESTQ-S has been shown to be an effective method of detecting early signs of overreaching, with the *Fatigue* and *Being in Shape* subscales significantly altered ($p < 0.05$) during periods of functional overreaching (Kellmann and Klaus-Dietrich, 2000; Coutts, Wallace and Slattery, 2007; Freitas *et al.*, 2014).

In order to increase the practical application of RESTQ-S in an applied setting, a modified, shortened version of the questionnaire has been developed (Kallus and Kellmann, 2016). Just under half the length, the shortened version (RESTQ-S-36) has 36 questions, divided into 12 subscales with 3 questions each (Nicolas *et al.*, 2019). A study of RESTQ-S-36 in 473 university students across various sports reported adequate model fit across male and female, elite and non-elite, team and individual-sport athletes (Nicolas *et al.*, 2019). The RESTQ-S-36 subscales fell within defined ranges for acceptable internal consistency, and second-order factor scores also demonstrated adequate reliability ($0.84 < \rho < 0.94$) (Nicolas *et al.*, 2019). Using RESTQ-S-36, a study of 72 national level swimmers found general recovery to be negatively correlated with training load, measured by session rating of perceived exertion (sRPE), during an overload period (r = -0.33) and taper period (r = -0.42), and total stress was positively correlated with training load during both overload (p = 0.40) and taper (r = 0.36) (Nicolas *et al.*, 2019). Although only moderate correlations, with factors outside of training impacting stress and recovery, a linear dose-response relationship is not to be expected, and these correlations do indicate the sensitivity of RESTQ-S-36 to changes in training load. Furthermore, a study of eleven swimmers during a 2-week taper phase found training load, again measured by sRPE, to be positively correlated with total stress (r = 0.58) and negatively correlated with total recovery (r = -0.53) (Nicolas *et al.*, 2019). Heart rate recovery in the minute following exercise was also reported to be moderately correlated with total recovery (r = 0.55) and total stress (r = -0.61), indicating that RESTQ-S-36 responses are not only associated with training load, but also correlated with objective recovery measures (Nicolas *et al.*, 2019).

With its sports-specific nature and subscales for both stress and recovery, RESTQ-S gained popularity quickly, estimated to have been disseminated to several thousand athletes in the first five years after its release (Davis, Orzeck and Keelan, 2007). RESTQ-S measures distinct features of stress and recovery, aiding coaches in identifying recovery-stress imbalance in athletes (Kallus and Kellmann, 2016). Additionally, it has been suggested that asking athletes about specific recovery activities, as done in RESTQ-S, increases awareness of recovery and results in improvements in athletes' recovery choices (Robson-Ansley, Gleeson and Ansley, 2009). However, despite its many benefits, there are several limitations of RESTQ-S that should be noted. Firstly, as mentioned previously, the structure and subscales have been questioned, with some subscales more responsive to changes in training load than others (Coutts, Wallace and Slattery, 2007; Saw, Main and Gastin, 2016). However, when training load is used for associations, the variation in the responsiveness of the subscales is to be expected, as some subscales are more closely related to off-pitch status rather than on-pitch stressors and recovery actions. Additionally, the length of the full-version of RESTQ-S can make it unwieldly for regular use in practical settings, especially when the original 3-day recall period is utilized (Nicolas *et al.*, 2019). Although, RESTQ-S-36 alleviates the issue of length, further studies will be needed to verify the effectiveness of this questionnaire for monitoring recovery-stress states, particularly in team-sport athletes. Therefore, until further research is performed, despite its length, RESTQ-S may be more appropriate than RESTQ-S-36 for monitoring recovery-stress state in team-sport athletes.

3.4.4 Acute Recovery Stress Scale (ARSS) and Short Recovery Stress Scale (SRSS)

The three recovery monitoring questionnaires discussed above, POMS, DALDA, and RESTQ-S, are the three most commonly studied subjective recovery measures in the literature (Saw, Main and Gastin, 2016). However, two relatively new questionnaires, the Acute Recovery and Stress Scale (ARSS) and the Short Recovery and Stress Scale (SRSS) also warrant consideration. Initially developed in Germany, these scales were designed to provide quick and simple methods of assessing the psychological and physiological components of recovery and stress (Nässi *et al.*, 2017a). Unlike other questionnaires which ask athletes to recall actions and feelings over a set period, ARSS and SRSS assess the current recovery-stress state of athletes by asking them to respond based on how they feel in a given moment (Nässi *et al.*, 2017b). Although beneficial in providing a snapshot of athletes' stress, recovery, and preparedness to train, this method of monitoring may be overly affected by events immediately prior to testing and provide an incomplete picture of overall wellbeing. ARSS consists of 32 adjectives, and athletes are asked to report on a 7-point Likert scale how much each adjective applies to them (Kölling *et al.*, 2015). Adjectives are then grouped and scores are calculated by combining the adjectives into 4 scales for stress and 4 for recovery, each made up of 4 adjectives (Kölling *et al.*, 2015). SRSS simplifies ARSS by pre-grouping the adjectives into their respective subscales with athletes only providing 8 responses, each based on 4 adjectives combined (Gabbett *et al.*, 2017). Both ARSS and SRSS were demonstrated to have satisfactory internal consistency (ARSS: $0.84 < \alpha < 0.96$, SRSS: $\alpha > 0.75$) (Nässi *et al.*, 2017a).

The German versions of SRSS and ARSS have been shown to be sensitive to changes in training load in a variety of sports and training (Kölling *et al.*, 2015; Wiewelhove *et al.*, 2015; Hammes *et al.*, 2016; Raeder *et al.*, 2016). Specifically, a study on 23 elite male cyclists found that overall stress and muscular stress, as measured by ARSS, significantly increased following a 6-day block of high-volume and high-intensity training ($p < 0.01$) and significantly decreased after a 3-day rest period ($p < 0.01$) (Hammes *et al.*, 2016). Similarly, SRSS scores for stress have been shown to notably increase and scores for recovery to notably decrease in response to a 6-day intensified strength training macrocycle in both male and female combat athletes and intermittent-ball sport athletes (Raeder *et al.*, 2016). Additionally, a study on the effect of high-intensity interval training in tennis players reported significantly increased perceived stress ($p <$

0.05) and decreased recovery ($p < 0.05$) as measured by SRSS following a 4-day shock microcycle (Wiewelhove *et al.*, 2015).

From a practical perspective, the largest challenge associated with using ARSS and SRSS in English-speaking athletes is that the original questionnaire is written in German. However, a recent study developed an English version of ARSS and SRSS, evaluating responses from 267 athletes and demonstrating satisfactory internal consistency for both questionnaires ($\alpha = 0.74$-0.89) (Nässi *et al.*, 2017a). However, more research will be needed to evaluate the effectiveness of the English translations of ARSS and SRSS. Additionally, another important limitation, particularly of SRSS, is that with only 8 questions, all the items and scales are clearly visible to the athlete, with the scoring method easy to decipher (Nässi *et al.*, 2017b). Therefore, there is a high risk for response bias in SRSS, and, if this questionnaire is used, efforts should be taken to ensure that athletes are answering honestly (Nässi *et al.*, 2017b). Overall, although more research is needed on the English translations, ARSS and SRSS show promise as methods of monitoring the current recovery and stress-state of athletes. However, it is important to note that by only assessing current state, ARSS and SRSS are distinct and provide unique information from other questionnaires that assess athlete wellbeing over a given period of time.

3.4.5 Subjective recovery monitoring in hockey

The previous sections demonstrate the breadth of research available on subjective recovery monitoring. Although the aforementioned questionnaires have been studied across a range of sports, there has been considerably less research on subjective recovery monitoring in hockey. Specifically, only three studies have reported athlete recovery as measured via validated recovery monitoring questionnaires in hockey populations, and all of these studies were performed over a short time domain during an international tour/tournament (Parrado *et al.*, 2010; Kölling *et al.*, 2015; Vescovi, 2019). Additionally, two studies have considered athlete wellness during international tournaments, but wellness responses have not been measured via a validated questionnaire (Ihsan *et al.*, 2017; McGuinness *et al.*, 2018). Specifically, athlete wellness (cumulative of score of fatigue, muscle soreness, mood state, and sleep quality, each ranked 0-10), was found to be significantly correlated with total distance ($r = -0.95$, $p = 0.004$), distance $\geq 15\ km \cdot h^{-1}$ ($r = -0.95$, $p = 0.003$), and distance $< 15\ km \cdot h^{-1}$ ($r = -0.94$, $p = 0.026$), normalized by minutes played and rating of perceived exertion (RPE) during six matches in a men's international hockey tournament (Ihsan *et al.*, 2017). Similarly, during seven matches in a women's international hockey tournament, a decrease in wellness (cumulative of score of muscle soreness, mood, and sleep quality, each ranked 0-10) was found to correspond to a decrease in running performance, and, in particular, a decrease in distance $\geq 16\ km \cdot h^{-1}$, although a correlational analysis was not performed (McGuinness *et al.*, 2018). Together these studies suggest a dose response relationship between running performance and athlete wellbeing in elite hockey during a tournament setting. However, as the questions used to assess wellness were not part of a validated questionnaire and the time domains considered were very short, it is not possible to know whether these wellness assessments provide a reliable indication of athlete recovery status and potential overreaching or overtraining.

Three studies have used validated questionnaires to evaluate athlete recovery in hockey, with mixed results (Parrado *et al.*, 2010; Kölling *et al.*, 2015; Vescovi, 2019). Firstly, a study conducted on eight Spanish male hockey athletes at the 2006 World Cup found that perceived tiredness scores, recorded via the French Society for Sports Medicine overtraining questionnaire (Brun, 2003), were strongly correlated with heart rate variability indices including the root mean square of differences of successive RR intervals (RMSSD) ($r = -0.73$) and the proportion of differences between adjacent RR intervals of more than 50 ms ($r = -0.81$) (Parrado *et al.*, 2010). These results suggested that perceived tiredness scores could provide an early indication of the autonomic response to overload during a competition setting. However, as no training load data were reported, these findings are notably limited in that only the relationship between subjective and objective recovery measures were considered, and the relationship with the amount of physical and physiological work performed is unknown. Overcoming this limitation, a study of the Canadian U21 hockey squad during a 16-day international tour, considered the relationship between athlete recovery measured via the Total Quality Recovery (TQR) questionnaire and wellness (cumulative score of fatigue, stress, sleep, muscle soreness, training enjoyment, irritability, and overall health, each ranked 0-10), and athlete training load (total distance, training impulse and RPE) (Vescovi, Klas and Mandic, 2019). No significant correlations were found for current day or next day analyses between Total Quality Recovery and training load, with the only significant correlation found occurring between wellness and total distance ($p = 0.038$) (Vescovi, Klas and Mandic, 2019). Although the relationship was significant, a one unit rise in wellness was found to only correspond to an increase in total distance of 37.8m, which is trivial from a practical perspective (Vescovi, Klas and Mandic, 2019).

Finally, the only study examining the use of recovery monitoring questionnaires discussed above in hockey was performed on the German Women's Junior National Team, evaluating the validity of ARSS and RESTQ-S (Kölling *et al.*, 2015). In this study, 25 athletes (aged 18-20) completed ARSS twice daily and RESTQ-S at the start and conclusion of a 5-day camp, which incorporated both training sessions and test matches (Kölling *et al.*, 2015). The results demonstrated that ARSS was sensitive to the increased training volume that athletes experienced during the camp, with overall recovery scores, measured by ARSS, significantly higher on the first two days of the camp than days 3 ($p < 0.030$), 4 ($p < 0.014$), 5 ($p < 0.003$) (Kölling *et al.*, 2015). Similarly, overall stress, measured by ARSS, significantly increased from the beginning to the end of the camp ($p < 0.001$), with differences also noted between day 2 and days 3 ($p < 0.001$), 4 ($p < 0.016$), and 5 ($p < 0.001$) (Kölling *et al.*, 2015). Individual ARSS profiles were also evaluated and one athlete who displayed clear signs of maladaptation throughout the camp was identified (Kölling *et al.*, 2015). Although only completed at the beginning and end of the training camp, RESTQ-S also showed sensitivity to the increased training demands, with scores for the subscales of *Physical Complaints* ($p < 0.007$) and *Injury* ($p < 0.001$) significantly increasing and *Physical Recovery* ($p < 0.047$) significantly decreasing (Kölling *et al.*, 2015). Overall, these findings indicate the sensitivity of both RESTQ-S and ARSS to changes in training load in hockey, demonstrating the potential usefulness of these questionnaires to monitor the recovery-stress balance and detect early signs of non-functional overreaching and overtraining in hockey athletes. However, it is worthwhile to note that this study was limited by its very short duration, which was not long enough to incur overtraining, and its sample of only female athletes (Kölling *et al.*, 2015). Additionally, as training load was not measured, the relationship between individuals' training loads and subjective responses could not be considered. Therefore, although this study set the foundation for using ARSS and RESTQ-S to monitor recovery and stress status in hockey athletes, further research is needed on the use of subjective recovery monitoring to assess athlete wellbeing and detect overreaching and overtraining in hockey athletes.

In summary, as discussed in this section, subjective recovery measures have been repeatedly shown to be valid methods of monitoring the balance of stress and recovery in athletes. POMS, DALDA, RESTQ-S, ARSS, and SRSS have been shown to be sensitive to changes in training load and to provide information to help identify early signs of overtraining. With the goal of recovery monitoring being the early detection of overreaching to ensure that non-functional overreaching and overtraining do not occur, the subjective recovery measures discussed in this section have been shown to be valid methods of monitoring athlete recovery. Based on these findings, for the research in this thesis, RESTQ-S was chosen as the subjective recovery measure of choice for several reasons. Firstly, its sports-specific nature and distinct information on both recovery and stress, rather than overall mood status, made it better suited than POMS. Additionally, as the research aimed to considered long-term relationships, with all subjects being volunteers, expecting athletes to complete lengthy questionnaires daily was not possible; therefore, DALDA was not chosen as a monitoring tool. Although DALDA has been used less frequently in some studies, it is also limited by a lack of research on team-sport athletes, and, with only three possible responses and no scoring system, there is limited potential for statistical analysis. Finally, although the study on ARSS and RESTQ-S in the German hockey team provided strong evidence in support of either of these questionnaires, RESTQ-S was selected over ARSS due to the lack of research on the English translation of ARSS and the ability of RESTQ-S to provide information on a set time-period rather than the current state of the athlete. Although RESTQ-S is somewhat lengthy and RESTQ-S-36 would have practical advantages, until more research was performed on this shortened-version of the questionnaire, particularly in team-sport athletes, using the full RESTQ-S was more prudent. Therefore, RESTQ-S was used to monitor the balance of recovery and stress and identify early signs of overreaching in the female hockey athletes in this study.

3.5 Objective recovery measures

In contrast to subjective recovery monitoring, objective recovery measures are based on non-perceptual tests of physiological function. Evaluating the physiological rather than the psychological aspects of recovery, objective monitoring encompasses measures of neuromuscular function, autonomic nervous system (ANS) activity, and blood-based markers of inflammation and immune response (Nässi *et al.*, 2017a; Coutts *et al.*, 2007; Johnston *et al.*, 2013; Chambers *et al.*, 1998; Andersson *et al.*, 2008; Starling *et al.*, 2019; Meeusen *et al.*, 2013; Bellenger *et al.*, 2016a; Cadegiani and Kater, 2017). Ideally, the best indicator of an athletes' physiological wellbeing and preparedness to compete is athletic performance (Saw, Main and Gastin, 2016). However, due to the strenuous and time-consuming nature of maximal performance tests, it is clearly impractical to perform maximal performance tests to assess athletes' recovery status daily, or even weekly (Saw, Main and Gastin, 2016). In fact, if an athlete is experiencing training distress, completing a maximal performance assessment is one of the most detrimental activities that can be performed, as the strain of the tests will only push the athlete further along the continuum towards overtraining syndrome (Meeusen *et al.*, 2013). As a result, the objective measures regularly used to assess athlete recovery are designed to provide a snapshot into the current recovery status of an athlete without causing additional physical strain or time-demands (Wehbe *et al.*, 2015). Specifically, it has been suggested that recovery measures for team-sport athletes be quick and simple to administer in multiple athletes, require minimal technology and technical expertise, be reliable, and not be highly strenuous (Fowles, 2006; Wehbe *et al.*, 2015). Several different types of objective measures, including jump, sprint, and cycle tests of neuromuscular fatigue, heart-rate based measures of ANS function, and blood-based markers for inflammation and immune system function have been considered as objective monitoring techniques (Nässi *et al.*, 2017a; Coutts *et al.*, 2007; Johnston *et al.*, 2013; Chambers *et al.*, 1998; Andersson *et al.*, 2008; Starling *et al.*, 2019; Meeusen *et al.*, 2013; Bellenger *et al.*, 2016a; Cadegiani and Kater, 2017).

Objective measures of athlete recovery have several key advantages over subjective measures, including a decreased risk of response bias, measurable information on physiological function, and ease of data analysis (Morgan *et al.*, 1987a; Rushall, 1990; Meeusen *et al.*, 2013; Kölling *et al.*, 2015; Kallus and Kellmann, 2016). Since objective measures do not ask athletes to self-report, there is a decreased risk of response distortion (Meeusen *et al.*, 2013; Kölling *et al.*, 2015). Although it is possible for athletes to intentionally perform poorly during physiological performance tests, such as CMJs, objective recovery measures are more difficult for athletes to fake than questionnaires on which athletes can easily report erroneous recovery levels (Meeusen *et al.*, 2013; Kölling *et al.*, 2015). Additionally, where subjective measures primarily provide information on psychological wellbeing and recovery-stress states, objective measures are designed to provide more detailed information on several physiological components of overtraining, including neuromuscular fatigue, inflammation, and immune system function (Saw, Main and Gastin, 2016). Furthermore, with the outcome of most objective measures being a single quantitative score, data analysis is simpler compared to the multiple subscales of POMS, DALDA, and RESTQ-S (Morgan *et al.*, 1987a; Rushall, 1990; Kallus and Kellmann, 2016).

Despite the many benefits of objective recovery monitoring, several important drawbacks and distinctions should be noted. Firstly, many methods of objective recovery monitoring require costly equipment and technical expertise (Meeusen *et al.*, 2013). Specifically, measures of ANS function require heart-rate monitors, and blood tests involve the use of analysis machines. For blood-based testing, coaches or sport scientists must also be trained in safe blood collection techniques and the inconvenience to the athletes of regular blood sampling must be considered. In addition to these practical considerations, one of the most important drawbacks of objective recovery monitoring is that these measures have been shown to be less sensitive to changes in recovery status and training load than subjective measures (Tobar, 2005; Saw, Main and Gastin, 2016; Nässi *et al.*, 2017b). As physiological symptoms of overreaching take longer to manifest and become measurable than changes in psychological state, objective measures often do not indicate overreaching until athletes have already progressed further along the overtraining continuum (Saw, Main and Gastin, 2016). Finally, there is crossover between objective recovery measures and fitness markers, with several measures frequently used in the assessments of both athlete fitness status and recovery (Buchheit, 2014; Daanen *et al.*, 2012; Schneider *et al.*, 2018). Although similar tests are utilized, athlete fitness and recovery status are two distinct, albeit related, constructs, and care should be taken to distinguish research on fitness and recovery markers. Overall, despite the drawbacks, several types of objective recovery measures have been evaluated as early indicators of overreaching and overtraining, and these measures will be discussed in the subsequent sections.

3.5.1 Neuromuscular fatigue

Objective recovery measures assess neuromuscular fatigue through a variety of physiological assessments including CMJs, long jumps, short-distance sprints, and cycle ergometer tests (Chambers *et al.*, 1998; Coutts *et al.*, 2007; Andersson *et al.*, 2008; Delextrat, Trochym and Calleja-Gonzalez, 2012; Johnston *et al.*, 2013; Wehbe *et al.*, 2015; Wiewelhove *et al.*, 2015; Nässi *et al.*, 2017a; Starling *et al.*, 2019). Exhaustive exercise paired with inadequate recovery results in muscle damage and neuromuscular fatigue (Freitas *et al.*, 2014). Depending on the extent of the physiological stress, neuromuscular fatigue may persist for extended periods of time causing decreased muscular power; therefore, persistent high levels of neuromuscular fatigue may indicate maladaptation (Nässi *et al.*, 2017a). Thus, short and simple tests of neuromuscular fatigue and muscular power are commonly used to regularly assess athlete recovery status (Chambers *et al.*, 1998; Coutts *et al.*, 2007; Andersson *et al.*, 2008; Delextrat, Trochym and Calleja-Gonzalez, 2012; Johnston *et al.*, 2013; Wehbe *et al.*, 2015; Wiewelhove *et al.*, 2015; Nässi *et al.*, 2017a; Starling *et al.*, 2019). Specifically, measures of lower-limb power, such as CMJs, short sprints, and cycle tests, are commonly monitored in lower-limb-dominant sports (Wehbe *et al.*, 2015). Single and repeated CMJs have both been shown to have high intra-day and inter-day reliability in Australian rules football athletes, particularly when mean force is measured (coefficient of variance - CV- single jump: 1.1%, 5 consecutive jumps: 2.4%) (Cormack *et al.*, 2008). Jump height during CMJs has been shown to have a slightly lower but still acceptable level of overall reliability with a CV of 5.2-5.9% (Cormack *et al.*, 2008).

A variety of maximal jump tests have been used to monitor athlete recovery and track fatigue in both team and individual sports (Nässi *et al.*, 2017a; Coutts *et al.*, 2007; Johnston *et al.*, 2013; Chambers *et al.*, 1998; Andersson *et al.*, 2008; Starling *et al.*, 2019; Delextrat, Trochym and Calleja-Gonzalez, 2012; Wiewelhove *et al.*, 2015). Specifically, CMJ height and jump efficiency in a multiple rebound jump test were shown to significantly decrease ($p < 0.05$) following a 6-day, running-based, high-intensity interval training program designed to induce overreaching in 11 male and 11 female team-sport athletes (Wiewelhove *et al.*, 2015). Significant increases in perceived muscle soreness and serum concentrations of creatine kinase (CK) ($p < 0.05$) were observed alongside the decreases in jump height and efficiency, suggesting that overreaching was achieved (Wiewelhove *et al.*, 2015). Additionally, jump height and efficiency were shown to return to baseline levels following 72 hours of recovery, indicating that CMJ height and jump efficiency are sensitive to acute changes in training load and recovery and an effective indicator of overreaching (Wiewelhove *et al.*, 2015). Other studies on ultra-endurance running (Chambers *et al.*, 1998), rugby (Coutts *et al.,* 2007; Johnston *et al.,* 2013) basketball (Delextrat, Trochym and Calleja-Gonzalez, 2012), and soccer (Andersson *et al.*, 2008), have also shown decreases in jump height, force, and efficiency following increases in training load. It has been suggested that horizontal jumps are a more sport-specific metric of lower-limb power for field-based sports than vertical jumps as these sports more often require force to be produced in the horizontal rather than the vertical plane (Dobbs *et al.*, 2015; Starling *et al.*, 2019). As a result, some studies have used standing long jumps or five-bound jumps for distance to assess neuromuscular fatigue (Coutts, Slattery and Wallace, 2007; Starling *et al.*, 2019). For example, in 16 male triathletes, five-bound scores decreased by 7.9% ($p < 0.05$) following a four-week overload training period during which DALDA stress scores increased and performance on a 3-mile time trial decreased ($p < 0.05$) (Coutts, Wallace and Slattery, 2007). These results suggest that like vertical jumps, horizontal jump tests are also sensitive to changes in training load (Coutts, Wallace and Slattery, 2007). In contrast with these findings, a controlled trial of volleyball athletes found no significant changes ($p > 0.05$) in CMJ height after an 11-day period of intensified training, despite significant changes in RESTQ-S scores ($p < 0.05$) (Nässi *et al.*, 2017a). However, the experimental group in this study included only 8 athletes, and the intensification of training that occurred over only 11 days may not have been adequate to induce the physiological symptoms associated with overreaching (Nässi *et al.*, 2017a).

Other measures of neuromuscular fatigue, not based on jumping, have also been evaluated as assessments of athlete recovery (Wehbe *et al.*, 2015; Wiewelhove *et al.*, 2015). A maximal effort 20 m sprint was used to monitor athlete fatigue levels following a running-based high-intensity interval training program, with running scores significantly increasing ($p < 0.05$) in alignment with significant decreases in CMJ height and increases in serum CK (Wiewelhove *et al.*, 2015). Similar decreases in 10 – 20 m sprints have also been noted following intense training in basketball ($p < 0.05$) (Delextrat, Trochym and Calleja-Gonzalez, 2012), match play in elite female soccer (Andersson *et al.*, 2008), and a 6-week period of deliberate overreaching in rugby (Coutts *et al.*, 2007). In all cases, scores improved after a recovery period, further supporting the use of short-sprints as a recovery measure (Coutts *et al.*, 2007; Andersson *et al.*, 2008; Delextrat, Trochym and Calleja-Gonzalez, 2012). Additionally, a study on 12 Australian rules football athletes examined peak-power during a 6-second sprint test on a cycle ergometer in order to monitor neuromuscular fatigue following match play (Wehbe *et al.*, 2015). Substantial decreases in peak power were reported 24 hours post-match compared to one-hour pre-match; however, when 90% confidence intervals for effect size were considered, the significance of these findings was unclear (Effect size: $= -0.40 \pm 0.41$) (Wehbe *et al.*, 2015). Although, cycle tests are not sport-specific for field-based sports, the authors of the study maintained that the concentric component of cycle tests provides important information not obtained from CMJs

(Wehbe *et al.*, 2015). Nevertheless, until more research is performed, it is unclear whether cycle tests are a valid method of measuring neuromuscular fatigue, particularly for detecting overreaching (Wehbe *et al.*, 2015).

Measures of neuromuscular fatigue have many practical benefits; however, it is important to ensure the proper application and interpretation of these assessments (Rousanoglou, Georgiadis and Boudolos, 2008; Dobbs *et al.*, 2015; Wehbe *et al.*, 2015; Starling *et al.*, 2019). Jumps and short sprints are very quick and relatively simple tests to administer (Wehbe *et al.*, 2015). This simplicity allows these assessments to be used in a variety of individual and team-sport settings, without lots of equipment or technical expertise required (Wehbe *et al.*, 2015). Athletes should be given adequate time to warmup prior to completing these types of assessments; however, this warmup can be combined with athletes' regular pre-training routine. If jump tests are used, it is important that baseline values are established over a period of time, as both horizontal and vertical jump scores are associated with muscular strength, which is typically not uniform across a group of athletes (Rousanoglou, Georgiadis and Boudolos, 2008; Dobbs *et al.*, 2015). Additionally, baseline testing is also important in ensuring that athletes are familiarized with the testing protocol, as an initial learning curve occurs when athletes begin jump testing (Starling *et al.*, 2019). Overall, when appropriate protocols are used and athletes are regularly monitored over time, jump and sprint-based assessments of neuromuscular fatigue provide valuable information on objective athlete recovery status (Nässi *et al.*, 2017a; Coutts *et al.*, 2007; Johnston *et al.*, 2013; Chambers *et al.*, 1998; Andersson *et al.*, 2008; Starling *et al.*, 2019; Delextrat, Trochym and Calleja-Gonzalez, 2012; Wiewelhove *et al.*, 2015).

3.5.2 Autonomic nervous system function

Autonomic nervous system (ANS) function, most frequently measured by heart rate variability and heart rate recovery, has also been suggested as a method of measuring athletes' physiological recovery status (Hedelin *et al.*, 2000; Bosquet *et al.*, 2003; Baumert *et al.*, 2006; Lambert and Borresen, 2006; Chalencon *et al.*, 2012; Dupuy *et al.*, 2013; Meeusen *et al.*, 2013; Nelson *et al.*, 2014; Aubry *et al.*, 2015; Thomson *et al.*, 2016a; Bellenger *et al.*, 2016a; Bellenger *et al.*, 2017). The autonomic nervous system is key in maintaining homeostasis during and after physical activity with cardiovascular function determined by the balance of parasympathetic and sympathetic modulation (Thomson *et al.*, 2016b). Critical in 'fight-or-flight' reactions and high-intensity activity, the sympathetic nervous system is dominant during exercise, with its reciprocal system, the parasympathetic nervous system primary during rest (McCorry, 2007; Daanen *et al.*, 2012). Since heart rate based measures provide insight into cardiac ANS status, monitoring heart rate during and after exercise is a non-invasive, practical method of assessing ANS function (Bellenger *et al.*, 2016a; Schneider *et al.*, 2018). Additionally, repeated exposure to exercise stress results in physiological adaptations, impacting the ANS response to exercise; therefore, assessing ANS function through heart rate monitoring is often utilized to track athlete fitness (Daanen *et al.*, 2012; Bellenger *et al.*, 2016a). In addition to fitness status, it has been suggested that ANS function, specifically sympathetic and parasympathetic modulation, may be impacted in cases of overreaching and overtraining, and, as a result, these measures have been examined as methods of monitoring athlete recovery status (Hedelin *et al.*, 2000; Bosquet *et al.*, 2003; Baumert *et al.*, 2006; Lambert and Borresen, 2006; Chalencon *et al.*, 2012; Dupuy *et al.*, 2013; Meeusen *et al.*, 2013; Nelson *et al.*, 2014; Aubry *et al.*, 2015; Thomson *et al.*, 2016a; Bellenger *et al.*, 2016a; Bellenger *et al.*, 2017).

Heart rate recovery (HRR), heart rate variability (HRV), and heart rate acceleration (HRA) have been evaluated as athlete recovery metrics with mixed results (Hedelin *et al.*, 2000; Bosquet *et al.*, 2003; Baumert *et al.*, 2006; Lambert and Borresen, 2006; Chalencon *et al.*, 2012; Dupuy *et al.*, 2013; Meeusen *et al.*, 2013; Nelson *et al.*, 2014; Aubry *et al.*, 2015; Thomson *et al.*, 2016a; Bellenger *et al.*, 2016a; Bellenger *et al.*, 2017). Heart rate acceleration measures the kinetics of heart rate increase at the onset of exercise, providing insight on the transition to sympathetic activation (Bellenger *et al.*, 2016a). Opposite of HRA, HRR, the rate of heart rate decline following exercise, provides information on the rate parasympathetic reactivation and sympathetic withdraw following exercise (Daanen *et al.*, 2012). Furthermore, heart rate variability, the variation in the length of time between successive heart beats, often measured post-exercise or at rest, monitors autonomic balance with increased HRV suggesting elevated parasympathetic activity relative to sympathetic activation (Meeusen *et al.*, 2013; Bellenger *et al.*, 2016a). According to the American College of Sports Medicine and European College of Sports Science's joint consensus statement on overtraining syndrome, although HRV is theoretically a useful measure for providing insight on cardiac autonomic status, the lack of consensus in the literature indicates that changes in HRV cannot be used to distinguish overreaching and overtrained athletes (Meeusen *et al.*, 2013). Similarly, in a systematic review, Bellenger *et al.* noted that although increases in post-exercise HRV and HRR were found in athletes displaying signs of overreaching and overtraining, these increases were also reported in athletes showing positive adaptation to training, demonstrating that these measures cannot distinguish maladapting athletes (Bellenger *et al.*, 2016a). Furthermore, resting HRV was found to be unaffected by overreaching (Bellenger *et al.*, 2016a).

The only heart rate based measure that has been shown to distinguish overreaching and overtrained athletes is HRA, with increases in HRA occurring in cases of positive adaptation to training and decreases shown in cases of training distress and non-functional overreaching (Bellenger *et al.*, 2016a). However, these conclusions were based on only two studies in which athletes showed signs of overreaching (Nelson *et al.*, 2014; Bellenger *et al.*, 2016b) and one study on athletes showing positive adaptation to training (Laffite *et al.*, 2003). Two other studies, not included in Bellenger *et al.*'s review, have also demonstrated HRA to be a valid indicator of acute exercise-induced fatigue (Thomson *et al.*, 2016b) and overreaching (Bellenger *et al.*, 2017). However, all of these studies have been performed in individual-sport endurance athletes, with no research on the relationship between training distress and HRA in team-sport athletes. Therefore, more research is needed to evaluate the validity and reliability of this method of recovery monitoring.

In summary, although HRA, HRV, and HRR provide information on cardiac ANS function and, in theory, would be useful tools for the early detection of overreaching and overtraining, a review of the literature suggests that HRV and HRR are not valid methods of identifying athletes experiencing training distress (Lambert and Borresen, 2006; Meeusen *et al.*, 2013; Bellenger *et al.*, 2016a). Although HRA shows promise as a recovery measure, with HRA increasing in athletes positively responding to training and decreasing in those with negative responses (Laffite *et al.*, 2003; Nelson *et al.*, 2014; Bellenger *et al.*, 2016b; Bellenger *et al.*, 2017), additional research is needed in team sport-athletes. From a practical perspective, heart-rate based measures are beneficial because they are non-invasive, and the results are often simple and quick to obtain (Schneider *et al.*, 2018). However, the variety of measurement techniques and the long list of outside factors that can impact heart rate (temperature, mood, stress, sleep, medication, tobacco, alcohol, etc) can limit the interpretation of these measures (Meeusen *et al.*, 2013; Schneider *et al.*, 2018). Additionally, in order to measure heart rate acceleration, an intense bout of exercise needs to be performed, which can exacerbate fatigue and increase training distress in overreaching athletes (Thomson *et al.*, 2016b). Overall, these results indicate that extreme care should be taken if heart rate based measures of ANS function are used to monitor athlete recovery, and other more accurate and reliable measures should be implemented when possible.

3.5.3 Blood-based measures

Although a variety of hormones, immunological markers, and other proteins have been studied as potential metrics of training distress, no single blood-based marker has been identified for the definitive diagnosis and early detection of overreaching and overtraining (Gleeson, 2002; Urhausen and Kindermann, 2002; Meeusen *et al.*, 2013; Cadegiani and Kater, 2017). The markers most regularly studied include CK, plasma glutamine, urea, blood lactate, ammonia, C-reactive protein, cortisol, growth hormone, and adrenocorticotropic hormone (Meeusen *et al.*, 2013; Cadegiani and Kater, 2017). However, several comprehensive reviews of the literature have indicated that none of these markers are valid methods of detecting overreaching or overtraining (Gleeson, 2002; Urhausen and Kindermann, 2002; Meeusen *et al.*, 2013; Cadegiani and Kater, 2017). Specifically, basal levels of hormones were found to be mostly normal in athletes displaying signs of overreaching and overtraining, and, where abnormal results were noted, differences were not consistent and significant across studies considered in a systematic review (Cadegiani and Kater, 2017). Additionally, with so many factors impacting most blood-based measures, including diet, sampling conditions, seasonal variation, menstrual status in female athletes, and time since exercise, it can be difficult to adequately control for confounding factors impacting blood-based measures (Meeusen *et al.*, 2013). As a result, the consensus among several review articles is that resting measures of blood-based markers are not valid tools for the detection of overreaching or overtraining (Gleeson, 2002; Urhausen and Kindermann, 2002; Meeusen *et al.*, 2013; Cadegiani and Kater, 2017).

The only case in which significant differences in blood-based measures have been found between athletes positively and negatively adapting to training is when blood samples are taken during maximal performance tests (Cadegiani and Kater, 2017). Specifically, maximal blood lactate concentration has been shown to decrease in overtrained athletes, while submaximal values either remain consistent or slightly decrease (Meeusen *et al.*, 2013). Additionally, growth hormone, prolactin, and adrenocorticotropic hormone were shown to have blunted acute responses to maximal exercise in instances of overreaching (Cadegiani and Kater, 2017). In both cases, these blood-based markers only provide valid information on training distress when measured during maximal exercise (Meeusen *et al.*, 2013; Cadegiani and Kater, 2017). However, as mentioned previously, it is not feasible or reasonable to regularly perform maximal performance tests to determine athletes' state, as these tests will only negatively contribute to athlete recovery and worsen cases of overreaching and overtraining (Meeusen *et al.*, 2013). Therefore, despite being scientifically valid, these measures have little practical application.

Although not directly useful in the detection of overtraining, one benefit of monitoring blood-based measures is in the exclusion of other possible causes of underperformance (Gleeson, 2002; Meeusen *et al.*, 2013). In many cases, overtraining is diagnosed via the exclusion of other potential causes of performance decrement, with overtraining syndrome sometimes being referred to as unexplained underperformance syndrome (Meeusen *et al.*, 2013). Blood-based markers provide information on other potential causes of underperformance, including viral infection, anemia, allergies, thyroid disorders, and other medical problems that could potentially interfere with recovery (Gleeson, 2002). Additionally, information from blood-based measures can be useful in determining the physiological stress and wellbeing of an athlete, which is important when prescribing training dose and determining readiness to train (Gleeson, 2002). For example, serum levels of C-reactive protein, a marker of acute and chronic inflammation, increase several thousand-fold in response to injury or infection, thereby providing information on athletes' overall health and readiness to train (Meeusen *et al.*, 2013; Souglis *et al.*, 2015b). However, despite smaller fluctuations in C-reactive protein occurring following competition (Souglis *et al.*, 2015a; Souglis *et al.*, 2015b), it is not be specific and responsive enough to provide a marker of overtraining (Singh *et al.*, 2011; Meeusen *et al.*, 2013; Wiewelhove *et al.*, 2015) However, in terms of recovery monitoring, no blood-based marker provides an early indication of training distress (Gleeson, 2002; Urhausen and Kindermann, 2002; Meeusen *et al.*, 2013; Cadegiani and Kater, 2017). Therefore, the usefulness of these measures is limited to cases where performance decrements have already occurred and the exclusion of other diagnoses is needed (Gleeson, 2002; Meeusen *et al.*, 2013). Overall, as there is significant cost and expertise required to collect and analyze blood samples, and the current evidence suggests that these samples are not able to provide definitive information on training distress, other measures are likely better suited for use in recovery monitoring (Gleeson, 2002; Urhausen and Kindermann, 2002; Meeusen *et al.*, 2013; Cadegiani and Kater, 2017).

3.6 Conclusion

Although the literature cited in this review provides good evidence on the effectiveness of various methods of recovery monitoring, it is important to consider the overall limitations across these studies. Specifically, although many studies stated that they examined athletes experiencing non-functional overreaching and overtraining, the monitoring period was usually not long enough to ensure that these states were achieved. Often the responsiveness of recovery measures to changes in training load and taper following a period of intense training has been examined, which does not distinguish functional and non-functional overreaching and cannot fully replace research on overtrained individuals.   There are ethical issues associated with intentionally inducing overtraining syndrome in athletes, so very little research has been performed on athletes displaying overtraining syndrome. Therefore, future research may benefit from a retrospective approach in which athletes are evaluated after a diagnosis of overtraining. Additionally, since the only known treatment for overtraining is rest, future studies could examine if any additional resources or treatments (psychological, physiological, etc) could help athletes increase the pace of recovery and prevent relapse in the future.

Monitoring athlete recovery is crucial to ensure that athletes positively respond to a given training stimulus and maintain balance between stress and recovery. When this balance is not maintained, long-term under-recovery leads to non-functional overreaching and overtraining, which result in extended periods of decreased performance paired with psychological and physiological distress. As extended rest is the only treatment for overtraining, the aim of recovery monitoring is to prevent overtraining by identifying when athletes are in a persistent state or recovery-stress imbalance. This research presented in the subsequent chapter considered CMJ height as well as RESTQ-S in hockey athletes. As discussed previously, RESTQ-S has been chosen for subjective monitoring for a variety of reasons, including its sports-specific nature, ability for weekly use, and the detail it provides on both recovery and stress state. To compensate for the weakness of both subjective and objective recovery measures and to provide a more complete perspective on athletes' recovery-stress status and overall wellbeing, multifaceted monitoring approaches including both subjective and objective measures are recommended (Saw, Main and Gastin, 2016; Duffield *et al.*, 2018). Therefore, in addition to RESTQ-S, CMJ height was also assessed as an objective recovery measure. The countermovement was selected because it has been repeatedly shown to be a valid and reliable indicator of neuromuscular fatigue and overreaching. Additionally, this assessment is simple, fast, and does not fatigue athletes. The research on ANS and blood-based measures have not clearly demonstrated that these metrics are valid methods of recovery monitoring. Thus, based on the findings of this literature review, RESTQ-S and CMJ height were selected as the recovery monitoring measures to be evaluated in hockey athletes.

3.7 Addendum – Impact of COVID-19

Following the completion of this literature review and prior to the commencement of data collection on recovery monitoring, the COVID-19 global pandemic occurred. As a result of this, significant changes were made to the planned structure and timing of the recovery monitoring study. Data collection was postponed and some of the intended variables were not able to be measured. Additionally, several participants had to drop out of the study due to contracting COVID-19 or being required to self-isolate, and as the UK went back into lockdown mid-way through the study, data collection was cut short. During the period just prior to and during the coronavirus pandemic there was also a notable increase in articles published on recovery monitoring in hockey. As these studies were not published when the recovery monitoring study was designed, the results of these studies were not taken into consideration in the methodological planning. Therefore, these studies have not been addressed in this literature review, which focused solely on the research available prior to late 2019. However, these studies will be considered below as well as the specific impacts of COVID-19 on the recovery monitoring data collection.

Despite only five studies evaluating recovery monitoring in hockey prior to late-2019 (Parrado *et al.*, 2010; Kölling *et al.*, 2015; Ihsan *et al.*, 2017; McGuinness *et al.*, 2018; Vescovi, 2019), eight additional studies were published on this topic by mid-2021 (Krueger *et al.*, 2019; Perrotta *et al.*, 2019a; Vescovi, 2019; Burt *et al.*, 2020; Tuft and Kavaliauskas, 2020; Walker *et al.*, 2020; González-Fimbres, Hernández-Cruz and Flatt, 2021; McMahon, Sharp and Kennedy, 2021). This notable increase was likely due to a combination of a significant rise in the monitoring of athlete-reported outcome measures (Jeffries *et al.*, 2020) and a push for many researchers to publish previous findings during the lockdown periods when no sport was occurring. Overall, these studies support the use of recovery monitoring in hockey athletes, although, as has been the case in other sports, no one best measure has emerged. Additionally, these studies reflect a shift towards objective recovery monitoring with all but one (Tuft and Kavaliauskas, 2020) considering at least one objective measure of athlete recovery.

Beginning with subjective measures of athlete recovery, four studies showed mixed results on the relationship between recovery measures and training load (Krueger *et al.*, 2019; Burt *et al.*, 2020; Tuft and Kavaliauskas, 2020; McMahon, Sharp and Kennedy, 2021). Specifically, a study on eleven male hockey athletes monitored over a six week in-season training period found no relationship (r = -0.046, -0.034; p = 0.336, 0.370) between training efficiency models (combining both internal and external training load) with daily cumulative wellness scores (sum of general stress, tiredness, and muscle soreness rated 1-5) (Tuft and Kavaliauskas, 2020). Also considering athlete wellness, a study of the Irish men's hockey team at the 2016 Olympics found that muscle soreness significantly correlated with total distance (r = 0.649, p = 0.031) and playerload (r = 0.630, p = 0.038); however, there were no correlations between the other wellness measures of stress and sleep quality with training load (McMahon, Sharp and Kennedy, 2021). Similarly, a study of eleven male hockey athletes found that reported athlete muscle soreness (reported on a 1-5 scale) was elevated both 1 hour and 24 hours following competition (Burt *et al.*, 2020). Therefore, there would appear to be a consensus on the increase in muscle soreness following hockey competition; however, self-reported muscle soreness is not a validated measure of athlete recovery, particularly for the early diagnosis of overreaching and overtraining, so the practical application and potential interpretation of these results is extremely limited (Jeffries *et al.*, 2020). Overcoming this limitation, a study of the German under-18 men's squad during a 5-day tournament measured stress and recovery via the Short Recovery and Stress Scale for Sport (SRSS), as well as asking athletes to self-report muscle soreness (Krueger *et al.*, 2019). Perceived recovery decreased over the course of the tournament ($p < 0.01$) while perceived stress ($p = 0.02$) and muscle soreness ($p < 0.01$) increased (Krueger *et al.*, 2019). When taken together the results of these studies would suggest that apart from muscle soreness, which is elevated following competition, general athlete wellness is not sensitive to changes in training load, and, until these wellness measures are standardized and validated, few interpretations can be made (Jeffries *et al.*, 2020). Additionally, the results of the SRSS questionnaire provide evidence that this tool is responsive to changes in training load and may be a good indicator of recovery status in hockey (Krueger *et al.*, 2019). However, as the study was only conducted over five days, more research will be needed to evaluate validated recovery questionnaires in hockey athletes over a longer time domain.

Moving on to objective measures of athlete recovery, several studies have found heart-rate based measures to be sensitive to changes in athlete recovery status (Perrotta *et al.*, 2019a; Vescovi, 2019; González-Fimbres, Hernández-Cruz and Flatt, 2021). Firstly, a study of the intra-individual variation in HRV found low CV <8.5% and percent standard error of the estimate (%SEE) ≤ 4.0% during a supine orthostatic test in elite male hockey athletes (Vescovi, 2019). The low variability suggests that supine HRV measures may be sufficiently sensitive to detect when true changes occur in the ANS (Vescovi, 2019). In addition, the study reported that variability was not improved during the fourth through sixth minute of the orthostatic challenge, suggesting that a three-minute supine test is a reliable indicator of ANS function and athlete recovery in hockey athletes (Vescovi, 2019). Shortening the testing protocol even further, a study of female youth hockey players during a four-week training camp reported strong agreement between an ultrashort (one-minute) and criterion (five-minute) seated measure of heart rate variability (interclass correlation coefficient- ICC = 0.979) (González-Fimbres, Hernández-Cruz and Flatt, 2021). Internal training load (TRIMP) was also measured as part of this study with changes in coefficient of variation of the logarithm of the root mean square of successive differences ($LnRMSSD_{CV}$) associated with TRIMP values ($p < 0.01$) (González-Fimbres, Hernández-Cruz and Flatt, 2021). Specifically, the researchers suggested that increases in internal training load should be used to develop fitness, provided LnRMSSD demonstrate positive athlete coping (González-Fimbres, Hernández-Cruz and Flatt, 2021). In agreement with these findings, a study of twelve international female hockey athletes reported a moderate correlation between $LnRMSSD_{CV}$ measured over the weekend and TRIMP during the prior week ($r = 0.40$) (Perrotta *et al.*, 2019a). Taken together, these studies on HRV would appear to support the use of this type of recovery monitoring in hockey athletes. Specifically, HRV is a reliable measure of ANS function even when shortened testing protocols are used or measurements are taken over the weekend, both of which support practical application. In addition, the association between TRIMP and $LnRMSSD_{CV}$ would suggest that HRV is sensitive to changes in training load. However, despite these promising findings, HRV is still extremely limited as a recovery monitoring measure as it has not been shown to differ between overreaching and overtrained athletes and athletes experiences positive adaptations (Meeusen *et al.*, 2013; Bellenger *et al.*, 2016a). Therefore, despite its reliability, practicality, and the association with TRIMP (Perrotta *et al.*, 2019a; González-Fimbres, Hernández-Cruz and Flatt, 2021) if the aim of recovery monitoring is to distinguish athletes who are entering a maladaptive training state, HRV is still not a useful measure.

In addition to HRV, several studies have investigated other objective measures of recovery in hockey athletes, including blood CK concentrations and CMJ height. In a study of eleven male hockey athletes, CK was elevated both 1-hour and 24-hours following a hockey match despite no changes in neuromuscular function, as measured via CMJ height and peak isokinetic knee extensor and flexor torque (Burt *et al.*, 2020). As CK increases despite maintenance of neuromuscular function, the authors suggest that increased CK is likely a result of changes in muscle energy processes not muscle membrane damage and is therefore not a valid indicator of exercise-induced muscle damage following hockey competition (Burt *et al.*, 2020). Similarly, a study of elite male under-18 hockey athletes reported increases in CK over the course of a training camp ($p < 0.01$) despite no significant changes in the neuromuscular fatigue measures of CMJ height and a repeated sprint assessment (Krueger *et al.*, 2019). In contrast with these findings, a study of the Canadian's men's national hockey team during the 2016 Olympics reported that post-match CK significantly correlated with total distance ($r = 0.55$) and playerload ($r = 0.41$), with the authors suggesting that these results were indicative of changes in neuromuscular function (McMahon, Sharp and Kennedy, 2021). However, as no direct measures of neuromuscular fatigue were considered, it is not possible to determine whether the increases in CK were associated with exercise-induced muscle damage. Finally, a study of a female university hockey team during their first four-week training block of the season reported an increase in CK across athletes, regardless of fitness level, while remaining within the 'normal' reference ranges for athletes (Walker *et al.*, 2020). Although training load and other blood-based measures of hormonal and hematological markers were considered, these results were not examined at the individual level and measurements were only taken at the start and end of the training block, so it is not clear whether specific increases in CK were indicative of individual athletes entering a state of overreaching or overtraining. Therefore, despite these studies on CK in hockey athletes, the validity of CK as a recovery measure is still not established. Specifically, although the evidence suggests that CK increases in response to increased training load in hockey, it is unclear whether this increase is associated with muscle damage and, consequently, if it is indicative of non-functional overreaching or overtraining syndrome.

Although there has been an increase into research on recovery monitoring in hockey, when examined together, the results of these studies provide little indication as to the best measures for monitoring recovery status in hockey athletes. In terms of subjective measures, the lack of standardization and validation of general wellness questionnaires limits the interpretation and practical application of any results. However, the research on the ARSS (Kölling *et al.*, 2015) and SRSS (Krueger *et al.*, 2019) provides early evidence in support of the use of these questionnaires for assessing athlete recovery-stress balance. The use of ARSS and SRSS is further supported by the results of a recent systematic review and meta-analysis of athlete-reported outcome measures that found that ARSS and SRSS outperformed other questionnaires in key measurement properties and that despite some limitations, there was good evidence to support their use to monitor athlete recovery (Jeffries *et al.*, 2020). However, as the studies measuring ARSS and SRSS in hockey populations have been very short in duration, more research will be needed over a longer time-domain to determine if these questionnaires can be used to identify athletes entering a maladaptive training state. Moving on to objective measures, the recent research in hockey is in alignment with the previous conclusions of a systematic review performed across sports which suggested that objective measures were not sensitive to changes in athlete recovery status (Saw, Main and Gastin, 2016). Specifically, in terms of HRV, the research in hockey would appear, at a surface level, to provide good evidence in support of the use of this recovery measure, with shorter measurement protocols having good reliability and moderate correlations with athlete training load (Perrotta *et al.*, 2019a; Vescovi, 2019; González-Fimbres, Hernández-Cruz and Flatt, 2021). However, none of these studies considered the ability of HRV markers to distinguish between athletes in positive and maladaptive training states, and multiple reviews of the literature have suggested that HRV does not differ in athletes experiencing overreaching and overtraining (Lambert and Borresen, 2006; Meeusen *et al.*, 2013; Bellenger *et al.*, 2016a). Similarly, although CK has been shown to increase following hockey competition and training, there appears to be no association between elevated CK and decreased neuromuscular function, suggesting that CK is not indicative of exercise-induced muscle damage and does not distinguish maladapting athletes (Walker *et al.*, 2020; McMahon, Sharp and Kennedy, 2021). As the key aim of recovery monitoring is differentiating between athletes responding well and poorly to a given training stimulus and changes in neither HRV nor CK levels provides an indication of this, both HRV and CK lack usefulness as athlete recovery measures when used in isolation. Therefore, future research into recovery monitoring in hockey should focus on ARSS and SRSS questionnaires, with the specific aim of determining if these questionnaires can be used to distinguish athletes entering a maladaptive training state.

Chapter 4: Match-Demands in Competitive Hockey - A Systematic Review and Meta-Analysis

The literature review in the previous chapter focused on the recovery aspect of athlete monitoring. As outlined, monitoring athlete recovery provides information on how athletes are responding to a given training stimulus and their recovery-stress status. However, the other key aspect of evaluating athletes' recovery-stress state is understanding the physical and physiological stressors that athletes experience by performing their sport. Thus, where the previous chapter focused on the response of athletes to a given training stimulus, this chapter will systematically review the physical and physiological demands of hockey. As preparing athletes for match performance is ultimately the goal of athletic training for hockey, knowledge of the match-demands is essential to designing an effective training process. The match-demands of hockey had not previously been reviewed in the literature, so the load data provided in this chapter provide key information for monitoring athletes within training programs to help coaches ensure they are prepared for the demands of competition.

4.1 Introduction

From a sports analysis perspective, hockey has several features that make it distinct from other field-based team sports (Reilly and Seaton, 1990; White and MacFarlane, 2013; Abbott, 2016; McGuinness *et al.*, 2017). Most importantly, hockey has unlimited rolling substitutions, meaning that athletes can interchange with players on the bench at almost any time (Abbott, 2016). These substitutions not only add a practical challenge in terms of analyzing individual athlete data, but also cause the demands of hockey to be more intermittent and varied than other invasion games (White and MacFarlane, 2013). The pitch size for hockey (55 m x 91.4 m) also differs from other field sports, and there are no offsides or restraining lines, so athlete movement patterns are stochastic and cannot easily be predicted (McGuinness *et al.*, 2017). Finally, hockey athletes must adopt a semi-crouched posture, when passing, receiving and dribbling, which has been shown to increase energy expenditure compared to normal running (Reilly and Seaton, 1990). Therefore, data from other field-based team sports cannot be accurately applied to hockey, and information on the demands of hockey must be derived from sport-specific research.

Measuring match-demands is a key aspect of athlete monitoring as it allows practitioners to accurately quantify the amount of work that athletes perform (Bourdon, 2017). Monitoring load and adjusting training dose accordingly has been shown to improve athlete performance and wellbeing, decrease injury occurrence, and increase athlete fitness in a range of intermittent ball sports (Foster *et al.*, 2001; Kevin and James, 2015; Mara *et al.*, 2015; Bourdon, 2017). The most effective form of training has been shown to be that which best mirrors the internal and external demands of competition, so monitoring training load during competition allows practitioners to have a template upon which to base training intensities (Gabbett, 2010; Liu *et al.*, 2013). In addition, athletes following identical training programs can perform varied external loads and have vastly different physiological responses to training. Therefore, monitoring individual training loads is crucial to ensuring that all athletes are receiving appropriate doses and are neither overtraining nor undertraining (Bourdon, 2017).

In alignment Impellizzeri *et al.'s* (2022) theoretical framework for the training process, the term training load will be used to describe internal and external load throughout this review. However, it is important to note that the data presented here are solely based on hockey competition not training. Training load can be measured in competition and although the term includes the word 'training,' this construct is not specific to a non-competitive environment (Impellizzeri, Marcora and Coutts, 2022). Therefore, training load measures are used and referenced throughout this review to summarize the physical and physiological match-demands of hockey.

Despite the importance of athlete monitoring and the increase in training load measurement in hockey over the past decade (Bourdon, 2017), no research has systematically reviewed the demands of hockey competition. Information on these demands allows for comparison of hockey with other sports, between hockey populations, and between training and competition. Describing the match-demands of hockey via training load also improves the understanding of the demands being placed on athletes and what preparatory and recovery protocols should to be in place around competition. Therefore, the aim of this current study was to determine match-demands of male and female hockey through a systematic review and meta-analysis of published studies in this field.

4.2 Methods

This study was performed in accordance with the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)* and calculations followed the guidance outlined in the *Cochrane Handbook for Systematic Reviews* (Moher *et al.*, 2009; Higgins *et al.*, 2019). The PRISMA checklist was used to guide the reporting of results, with specific consideration to the risk of bia both within and across studies.

4.2.1 Eligibility criteria

Observational studies written in the English language and published in peer-reviewed journals were eligible for inclusion in this review. To be included, studies must have measured training load in outdoor hockey, in a competition setting, with data from at least one complete match included. Training load was defined to be any measure of the work performed during a session summarized as a numerical score, with internal and external training load measures both included in this analysis. To provide a comprehensive review of the sport, no exclusions were made based upon age, sex or performance level. However, articles focusing on indoor hockey were excluded due to the vast differences in the rules of indoor and outdoor hockey. Additionally, studies only reporting training data or incomplete competition data (not reported across entire matches) were excluded. In 2009, the International Hockey Federation updated the rules of hockey, introducing the self-pass, which has been shown to significantly increase the pace and intensity of the game (Tromp and Holmes, 2011). Therefore, to make the results of this review most relevant to the game of hockey as it is currently played, only studies in which data collection occurred after the 2009 rule change were considered. Studies performed across multiple sports were eligible for inclusion, except in the case where only pooled data were presented and hockey-specific results could not be extracted. Articles consisting of abstracts only and review articles were excluded.

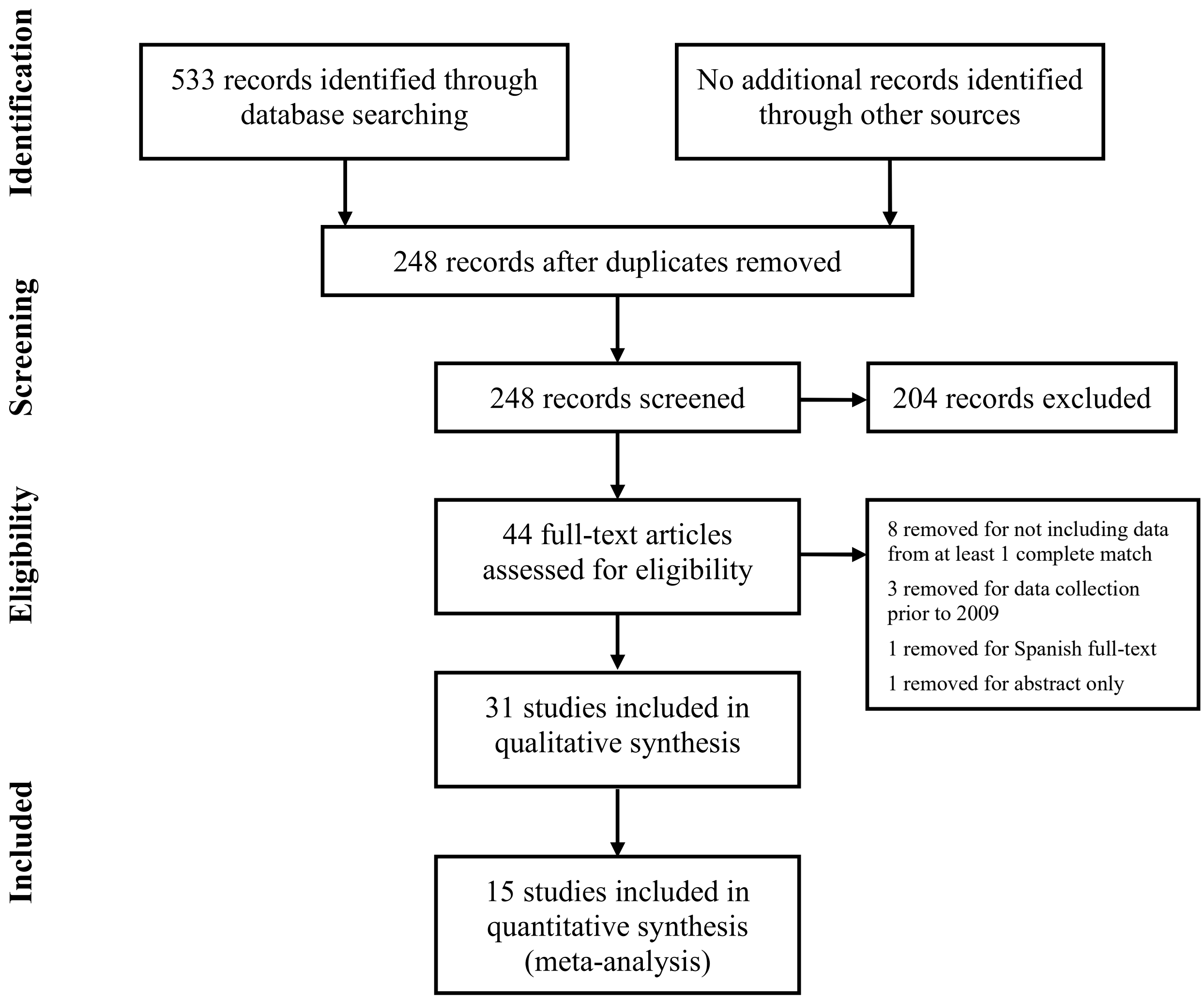


Figure 4.1: Study Selection Flowchart

4.2.2 Information sources

An electronic search was performed of MEDLINE, Scopus, Science Direct, SPORTDiscus, and Web of Science. Dates were restricted to exclude articles published before 2009. The final search was conducted on 12/01/2020. The decision was made not to update this literature review to include hockey after this date due to the occurrence of the COVID-19 pandemic and the severe disruption caused in domestic and international hockey. International hockey was suspended during the pandemic, and, due to the negative impact on training during this time, the study was also not updated to include data collected immediately following the pandemic. Therefore, the period of hockey considered was that from the beginning of the self-start rule to the start of the COVID-19 pandemic and the subsequent interruption in hockey. The search terms were as follows: (Hockey NOT 'Ice Hockey') AND ('training load' OR GPS OR 'global positioning system' OR 'heart rate' OR 'training impulse' OR TRIMP OR distance OR 'physical demands' OR 'physiological demands' OR workload).

4.2.3 Study selection

The researcher screened all studies identified to determine those that met the eligibility criteria. After the removal of duplicate studies, abstract and title screening was performed, followed by full-text screening. After initial database searching, 535 potential records were identified (Web of Science - 177, Scopus - 138, SPORTDiscus - 111, Medline - 77, Science Direct - 32). Following the removal of duplicates and screening for eligibility criteria, 31 records remained. A schematic of study selection is displayed in Figure 4.1.

4.2.4 Data collection process, data items and summary measures

Data collection was performed independently by the researcher. Sample size, sex, number of matches monitored, and performance level were recorded for each study. Additionally, the following data were extracted (1) training load variables measured, (2) monitoring equipment used, (3) monitoring setting (ie, tournament, league, test series), (4) participant information, (5) all training load data reported, (6) positional, performance level, or other distinctions made in data reported, (7) data analysis techniques used, and (8) statistical tests performed. When provided, any additional information of note regarding participants, setting, data collection and analysis was also recorded. The summary measures considered were training load variables presented as means ± standard deviations.

The study:

1. describes the setting and/or participating locations.
2. describes relevant dates (exposure, follow-up, data collection, etc.).
3. provides statement concerning institutional review board (or equivalent) approval and participant consent.
4. defines inclusion and exclusion criteria, including the health and injury history of participants.
5. adequately defines training load variables.
6. adequately defines data acquisition methodology / technology.
7. describes relevant team competition level.
8. provides details on player positions.
9. collected data over multiple exposures.
10. reports reliability of measurement methodology.
11. collected data on more than one team.

Figure 5.2 Quality Assessment Checklist (Taylor et al., 2017)

4.2.5 Risk of bias

Quality appraisal was performed on all studies according to a modified version of the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for the assessment of methodical quality. This checklist was selected because it was designed for the assessment of observational studies and provides clear guidance on the items to be assessed (Von Elm *et al.*, 2007). Additionally, it has been used in published systematic reviews on injuries (McKeon and McKeon, 2012; Doherty *et al.*, 2014), training design (Clark, McEwan and Christie, 2019), performance (Altmann *et al.*, 2015; Tawa and Louw, 2018), and the activity demands across several team sports (Taylor *et al.*, 2017). As was performed in previous systematic reviews (McKeon and McKeon, 2012; Doherty *et al.*, 2014; Taylor *et al.*, 2017), the STROBE checklist was modified to ensure that the questions for quality appraisal were appropriate to the research question. Due to the similarities of the research question and data items, the modification selected for this study was adapted from that given by Taylor *et al.*, in their systematic review of activity demands during multi-directional team sports (Taylor *et al.*, 2017). This modification consists of 11 items, given in Figure 4.2, each with a yes (1) or no (0) response for a maximum possible score of 11 (Taylor *et al.*, 2017).

4.2.6 Synthesis of results

Meta-analyses were performed on total distance and workrate data separately for each sex. These analyses were performed both overall, for all playing positions combined, and separately by position (forward, midfield, defense). Although studies on athletes competing at all levels were included in the systematic review, only data collected on elite hockey were eligible for inclusion in the meta-analysis. This distinction was made in order to provide a more homogenous group for analysis and create greater specificity in the output. A similar analysis would have been performed on sub-elite hockey had there been sufficient data available. Elite hockey was defined as international and junior international competitions as well as matches in semi-professional leagues (Australian Hockey League and England Hockey Premier Division). One study reported data on both sub-elite and elite athletes, so only data from the elite athlete subset were incorporated in the meta-analysis (Buglione *et al.*, 2013). Studies presenting data on total distance and workrate in elite hockey were excluded from either positional or overall analyses for providing summary data by position rather than by individual (Jennings *et al.*, 2012c) or including time on bench in workrate data (White and Macfarlane, 2015a; Morencos *et al.*, 2019).

Data were analyzed in Microsoft Excel (Microsoft Corporation, Version 2002, Redmond, Washington), via an inverse variance random effects model, as described by Hedges and Vevea (Hedges and Vevea, 1998). A sample of the calculations for overall total distance are provided in Appendix B. For the purpose of these meta-analyses, the sample size was taken to be the number of athletes participating in the study, rather than the number of match-files analyzed. This distinction was made to avoid disproportionate sample allocations due to the repeated measures data and to minimize the impact of within-athlete correlations, which were ignored in some studies (Bland and Altman, 1994). When positional data were provided without positional breakdowns in sample size, the sample size for each position was determined either from the positional breakdown of match-files (Polglaze *et al.*, 2015), or based on the assumption of a uniform distribution across positions (McGuinness *et al.*, 2019). Results are presented as means ± standard deviations, with 95% confidence intervals shown in the forest plots. In accordance with the work of the Hedges and Vevea (1998), significance tests were performed corresponding to the confidence intervals constructed using the inverse variance random effects model. Thus, following the meta-analyses, overall results were compared across sexes using independent samples t-tests. Additionally, results were compared across positions using one-way analysis of variance (ANOVA), with Tukey's honestly significant difference (HSD) post-hoc test.

Due to the variety of training load outcomes reported and hockey populations studied, meta-analyses were not performed on other training load data. However, for consistency, some data were combined across groups (for example when data from attacking midfielders and defensive midfielders were presented separately), using pooled standard deviations (Higgins *et al.*, 2019). Additionally, when distance data were provided separately across halves or quarters of match-play, these data were summed, with a perfect correlation assumed for the calculation of standard deviation. This method of calculation provides a conservative estimate of variance, taking into consideration the lack of independence of individuals' work across various portions of match-play. Within individual studies, when comparisons were not reported across positional groups, but sufficient data were available, comparisons were made via one-way ANOVA with Tukey's HSD test. Significance was set at $p < 0.05$.

4.3 Results

Table 4.1: Study Characteristics and Participant Information

Authors/Date	Sex	Athletes	Matches	Level	Elite	Age (y)	Height (cm)	Mass (kg)	VO2 (ml·kg^{-1}·min^{-1})	GPS±GNSS	HR	Match Setting
Gabbett 2010	F	14	32	N	E	23.3 ± 3.2			53.5 ± 4.3	Catapult MinimaxX 5 Hz		L
Lythe & Kilding 2011	M	18	5	I	E	24.4 ± 4.5	180.1 ± 4.9	78.4 ± 6.5	64.9 ± 1.9	GPSports SPI Elite 1 Hz	Polar Team	TS
Jennings *et al.* 2012a	M	15	6 (90)	I	E	27 ± 4	179 ± 5	77 ± 5	64.2 ± 3.1	Catapult MinimaxX 5 Hz		T
Jennings *et al.* 2012b	M	31	14 (224)	I & N	E	National: 22 ± 4 International: 27 ± 4	National: 178 ± 8, International: 179 ± 5	National: 78 ± 9 International: 77 ± 5		Catapult MinimaxX Team 2.5 5Hz		T & T
Buglione *et al.* 2013	M	22	6 (66)	I & N	E & S	International: 25.2 ± 3.8, National: 22.2 ± 4.0	International- 174 ± 5.2, National- 172 ± 1.8	International: 70.3 ± 2.6 National: 66.7 ± 6.2	International: 55.8 ± 2.8, National: 52.0 ± 2.5	GPSports SPI Elite 1 Hz	Polar HR Strap	T & L
Liu *et al.* 2013	M	38	24 (38)	N	S							T
Lythe & Kilding 2013	M	18	5 (21)	I	E	24 ± 4.5	180.1 ± 4.9	78.4 ± 6.5		GPSports SPI Elite 1 Hz	Polar Team	TS
White & MacFarlane 2013	M	16	8 (73)	I	E	25 ± 4		70.9 ± 6.6	61 ± 2.1	Catapult MinimaxX 5 Hz		T
Vescovi 2014	F	44	8 (123)	JI	E					GPSports SPI Pro 5 Hz		TS
Polglaze *et al.* 2015	M	24	7 (105)	I	E	27.0 ± 2.7		78.8 ± 6.8	62.0 ± 3.8	Catapult MinimaxX S4 10 Hz		T
Vescovi & Frayne 2015	F	68	6 (68)	N	S					GPSports SPI Pro 5 Hz		L
White & MacFarlane 2015a	M	16	8 (75)	I	E	25 ± 4		70.9 ± 6.6	61.0 ± 2.1	Catapult MinimaxX 5 Hz		T
White & MacFarlane 2015b	F	108	(186)	N	S	Range:16-39				Catapult MinimaxX 5 Hz		L
Kim *et al.* 2016	F	32	20	I	E	28.2 ± 3.1	164.9 ± 4.2	59.6 ± 4.2		GPSports SPI-HPU 5 Hz		T & TS
Vescovi 2016	F	44	8	JI	E	U17 & U21	165.5 ± 6.0,	60.6 ± 8.8,		GPSports SPI Pro 5 Hz	Polar HR Strap	TS
Crewther *et al.* 2017	F	23	4 (62)	I	E	25.6 ± 3.6	168.1 ± 8.2	62.8 ± 6.9				TS

Authors/Date	Sex	Athletes	Matches	Level	Elite	Age (y)	Height (cm)	Mass (kg)	VO2 ($ml \cdot kg^{-1} \cdot min^{-1}$)	GPS±GNSS	HR	Match Setting
Ihsan *et al.* 2017	M	12	6 (72)	I	E	22.3 ± 1.5	175 ± 2.5	73.5 ± 8.2		Catapult MinimaxX Team 2.5 5 Hz		T
Perrotta *et al.* 2017	F	16	4 (60-63)	I	E	22.0 ± 2.1	169.7 ± 3.5	61.3 ± 5.7	53.5 ± 7.9		Polar Team[2]	TS
Sunderland & Edwards 2017	M	20	17 (234)	N	E	21.1 ± 3.4	178 ± 6	73.8 ± 8.2		GPSports SPI Elite 1 Hz		L
Casamichana *et al.* 2018	M	16	17 (145-146)	N	S	25.5 ± 2.9	177 ± 5	74.6 ± 5.5		GPSports SPI Elite 10 Hz		L
Chesher *et al.* 2018	M	15	6 (90)	I	E	27.3 ± 8.5	179.6 ± 6.0	75.9 ± 6.7		Catapult MinimaxX S4 10 Hz		T
McGuinness *et al.* 2018	F	27	15 (154)	I	E	23 ± 3	162.6 ± 13.0	66.0 ± 6.0		Visuallex Sport VX110 Log 10 Hz	Firstbeat version 4.5.0.2	TS
Morencos *et al.* 2018	M	16	17 (113)	N	S	25.5 ± 2.9	177.1 ± 5.3	74.6 ± 5.5		GPSports SPI Elite 10 Hz		L
Perrotta & Warburton 2018	F	17	4 (64)	I	E	22.0 ± 2.1	167.4 ± 5.3	62.5 ± 5.5	51.8 ± 2.8		Polar Team[2]	TS
Polglaze *et al.* 2018	M	16	6 (92)	I	E	27.5 ± 3.1		77.0 ± 6.2	63.3 ± 2.4	Catapult MinimaxX S4 10 Hz		T
Vescovi & Klas 2018	F	14	8 (96 GPS, 48 HR)	JI	E	18.8 ± 1.2	165.9 ± 6.3	64.6 ± 9.3		GPSports SPI Pro 5 Hz	Polar T31	TS
Vinson *et al.* 2018	F	13	18 (204)	N	E	28.0 ± 7.0				GPSports SPI HPU 15 Hz		L
Krueger *et al.* 2019	M	18	3 (48)	JI	E	16.6 ± 0.6	182.1 ± 5.5	73.8 ± 7.8		Catapult OptimEye S5 10 Hz	Acentas GmbH	T
McMahon & Kennedy 2019	F	19	25 (400)	I	E	23 ± 4		63.6 ± 5.5	57.5 ± 6	Catapult OptimEye 10 Hz		T & TS
Morencos *et al.* 2019	F	16	5 (50)	I	E	24.7 ± 2.8	165.2 ± 4.9	57.9 ± 5.9		GPSport SPI Elite 10 Hz		T
Vescovi *et al.* 2019	F	16	8 (112)	JI	E	18.8 ± 1.2	165.9 ± 6.3	64.6 ± 9.3		GPSports SPI Pro 5 Hz	Polar HR Strap	TS

* Number in parenthesis is the total number of player-match datafiles, as not all participants were monitored in all matches in some studies.
I: International, N: National, JI: Junior International, E: Elite, S: Sub-elite T: Tournament, L: League, TS: Test Series

Table 4.2: Risk of Bias

Authors	**QAC1**	**QAC2**	**QAC3**	**QAC4**	**QAC5**	**QAC6**	**QAC7**	**QAC8**	**QAC9**	**QAC10**	**QAC11**	**Total:**
Gabbett 2010	1	0	1	1	1	1	1	1	1	0	0	8
Lythe & Kilding 2011	1	1	1	0	1	1	1	1	1	1	0	9
Jennings *et al.* 2012a	1	1	1	0	1	1	1	1	1	1	0	9
Jennings *et al.* 2012b	1	1	1	0	1	1	1	1	1	0	1	9
Buglione *et al.* 2013	1	1	1	0	1	1	1	0	1	1	1	9
Liu *et al.* 2013	0	1	1	0	1	1	1	1	0	1	1	8
Lythe & Kilding 2013	1	0	1	0	1	1	1	1	1	1	0	8
White & MacFarlane 2013	1	1	1	1	1	1	1	0	1	1	0	9
Vescovi 2014	0	1	1	1	1	1	1	1	1	1	1	10
Polglaze *et al.* 2015	1	1	1	0	1	1	1	1	1	1	0	9
Vescovi & Frayne 2015	1	1	1	0	1	1	1	1	0	1	1	9
White & MacFarlane 2015a	1	0	1	1	1	1	1	0	1	1	0	8
White & MacFarlane 2015b	1	1	1	1	1	1	1	0	1	1	1	10
Kim *et al.* 2016	0	1	1	1	1	1	1	1	1	1	0	9
Vescovi 2016	0	1	1	0	1	1	1	1	1	1	1	9
Crewther *et al.* 2017	1	1	1	0	1	1	1	0	1	1	0	8
Ihsan *et al.* 2017	1	1	1	0	1	1	1	0	1	1	0	8
Perrotta *et al.* 2017	1	1	1	0	0	1	1	0	1	0	0	6
Sunderland & Edwards 2017	1	1	1	0	1	1	1	1	1	1	0	9
Casamichana *et al.* 2018	1	1	1	0	1	1	1	1	1	1	0	9
Chesher *et al.* 2018	1	1	1	0	1	1	1	1	1	1	0	9
McGuinness *et al.* 2018	1	1	1	1	0	1	1	1	1	1	0	9
Morencos *et al.* 2018	0	1	1	0	1	1	1	1	1	0	1	8
Perrotta & Warburton 2018	0	1	1	0	1	1	1	0	1	0	0	6
Polglaze *et al.* 2018	1	1	1	0	1	1	1	1	1	1	0	9
Vescovi & Klas 2018	1	1	1	0	1	1	1	0	1	1	0	8
Vinson *et al.* 2018	1	1	1	0	1	1	1	1	1	0	0	8
Krueger *et al.* 2019	1	1	1	1	1	1	1	0	1	1	0	9
McMahon & Kennedy 2019	1	1	1	1	1	1	1	1	1	1	0	10
Morencos *et al.* 2019	1	1	1	0	1	1	1	1	1	1	0	9
Vescovi *et al.* 2019	1	1	1	0	1	1	1	0	1	1	0	8

Quality assessment scores based on the Quality Assessment Checklist (Taylor et al., 2017)

Table 4.3: Training Load Measures Reported

Authors	**TD**	**Speed Zones**	**Max Speed**	**Accel /Decel**	**Player-load**	**Metabolic Power**	**Minutes**	**Workrate**	**Heart Rate**	**sRPE**	**TRIMP**	**Distinctions**	
												Positional	**Half/Qtr**
Gabbett 2010	✓	✓		✓								✓	
Lythe & Kilding 2011	✓	✓					✓		✓			✓	✓
Jennings *et al.* 2012a	✓	✓					✓					✓	
Jennings *et al.* 2012b	✓	✓						✓				✓	✓
Buglione *et al.* 2013	✓	✓		✓			✓		✓				✓
Liu *et al.* 2013	✓	✓										✓	✓
Lythe & Kilding 2013	✓	✓					✓		✓			✓	✓
White & MacFarlane 2013	✓	✓	✓	✓	✓			✓					
Vescovi 2014			✓									✓	
Polglaze *et al.* 2015	✓				✓		✓	✓				✓	
Vescovi & Frayne 2015	✓	✓	✓	✓		✓	✓	✓				✓	✓
White & MacFarlane 2015a	✓	✓	✓	✓	✓		✓	✓					
White & MacFarlane 2015b	✓	✓		✓	✓			✓					
Kim *et al.* 2016	✓	✓	✓									✓	
Vescovi 2016	✓	✓				✓	✓	✓	✓		✓	✓	
Crewther *et al.* 2017										✓			
Ihsan *et al.* 2017	✓	✓		✓			✓			✓			
Perrotta *et al.* 2017										✓	✓		
Sunderland & Edwards 2017	✓	✓					✓	✓				✓	
Casamichana *et al.* 2018		✓	✓				✓					✓	
Chesher *et al.* 2018				✓								✓	✓
McGuinness *et al.* 2018	✓	✓	✓				✓	✓	✓			✓	✓
Morencos *et al.* 2018	✓	✓	✓	✓			✓	✓				✓	✓
Perrotta & Warburton 2018									✓	✓	✓		
Polglaze *et al.* 2018	✓	✓		✓		✓	✓	✓				✓	✓
Vescovi & Klas 2018	✓	✓							✓				
Vinson *et al.* 2018	✓	✓	✓				✓	✓				✓	
Krueger *et al.* 2019	✓	✓		✓			✓	✓	✓	✓			
McMahon & Kennedy 2019	✓	✓			✓		✓	✓				✓	
Morencos *et al.* 2019	✓	✓		✓			✓					✓	✓
Vescovi *et al.* 2019	✓									✓	✓		

Table 4.4: External Training Load

Authors	**Total Distance (m)**				**Workrate (m·min^{-1})**				**Maximal Sprint Speed (m·s^{-1})**			
	All	**Defense**	**Midfield**	**Forward**	**All**	**Defense**	**Midfield**	**Forward**	**All**	**Defense**	**Midfield**	**Forward**
Gabbett 2010		6643 ± 1618	6931 ± 1882	6154 ± 271								
Lythe & Kilding 2011	6798 ± 2009 8160 ± 428								26.8			
Jennings et al. 2012a	9775 ± 618	9453^ ± 579	10160^ ± 215	9819 ± 720								
Jennings et al. 2012b	I: 9776 ± 720 N: 8589 ± 623											
Buglione et al. 2013	E: 7062 ± 1015 S: 6186 ± 997											
Liu et al. 2013	7334 ± 877	6671*^ ± 745	7733^ ± 729	7709* ± 720								
Lythe & Kilding 2013				8434 ± 301								
White & MacFarlane 2013	5819 ± 687				124 ± 17				27.3 ± 1.6			
Vescovi 2014									24.4 ± 1.6	24.1 ± 1.2	24.4 ± 1.4	24.7 ± 2.0
Polglaze et. Al 2015	6095 ± 938	6257* ± 909	6256` ± 931	5409*` ± 689	131 ± 11	120^* ± 8	136^` ± 10	129*` ± 9				
Vescovi & Frayne 2015	6493 ± 1334	6556 ± 1120	6765 ± 1392	6062 ± 1371	106 ± 12	98*^ ± 11	109^ ± 11	110* ± 11	24.4 ± 1.6	23.8 ± 2.0	24.4 ± 1.4	25.0 ± 1.4
White & MacFarlane 2015a	5868 ± 665				78 ± 11							
White & MacFarlane 2015b												
Kim et al. 2016	5270 ± 644	5383	5736	4815					27.1 ± 1.1	26.7	27.5	27.2
Vescovi 2016	4351 ± 1282	5143* ± 759	4735` ± 1305	3283*` ± 842	109 ± 8	103*^ ± 9	113^ ± 6	111* ± 6				
Crewther et al 2017												
Ihsan et al. 2017	5232 ± 479				126 ± 5							
Perrotta et al. 2017												
Sunderland & Edwards 2017	6603 ± 1089	7116* ± 1243	6811` ± 778	5881`* ± 774	139 ± 15	125^* ± 12	140^`± 5	151*` ± 13				
Casamichana et al. 2018												
Chesher et al. 2018												
McGuinness et al. 2018	4847 ± 583	5181*^ ± 607	4740^ ± 530	4549* ± 546	128 ± 15	115* ± 14	132` ± 15	142*` ± 17				

Authors	Total Distance (m)				Workrate ($m \cdot min^{-1}$)				Maximal Sprint Speed ($m \cdot s^{-1}$)			
	All	Defense	Midfield	Forward	All	Defense	Midfield	Forward	All	Defense	Midfield	Forward
Morencos et al. 2018	7675 ± 1065	7533 ± 968	7867 ± 1140	7611 ± 1058	111 ± 15	106^ ± 12	116^ ± 15	110 ± 14				
Perrotta & Warburton 2018												
Polglaze et al 2018	5523 ± 632	5856* ± 672	5556` ± 457	5083*` ± 479	117 ± 11	114^ ± 11	123`^ ± 9	114` ± 10				
Vescovi & Klas 2018												
Vinson et al. 2018					111 ± 9	100^ ± 4	117^ ± 5	113 ± 7				
Krueger et al. 2019	5839 ± 997				115 ± 10							
McMahon & Kennedy 2019	5029 ± 995	5206* ± 1067	5318` ± 870	4561*` ± 914	111 ± 13	105*^ ± 14	116^ ± 12	116* ± 13				
Morencos et al 2019	5687 ± 905				72 ± 10							
Vescovi et al. 2019												

Significance difference ($p < 0.05$) between: *forward and defense, ^midfield and defense, `forward and midfield. Grey highlighting indicates total distance data by position not per athlete and workrate data includes time on the bench. I: International, N: National, E: Elite, S: Subelite.

Table 4.5: Distance in Speed Zones (m)

Authors		**Zone 1**	**Zone 2**	**Zone 3**	**Zone 4**	**Zone 5**	**Zone 6**
Gabbett 2010		0-3.6 km·h^{-1}	3.6-10.8 km·h^{-1}	10.8-18 km·h^{-1}	18-25.2 km·h^{-1}	>25.2 km·h^{-1}	
	D	841 ± 229	3618 ± 821	1763 ± 566	369 ± 178	52 ± 62	
	M	681 ± 243	3422 ± 989	2181 ± 558	571 ± 244	77 ± 69	
	F	728 ± 231	3017 ± 247	1941 ± 198	423 ± 195	46 ± 57	
Lythe & Kilding 2011		0-6 km·h^{-1}	6.1-11 km·h^{-1}	11.1-14 km·h^{-1}	14.1-19 km·h^{-1}	19.1-23 km·h^{-1}	>23 km·h^{-1}
	All	2410 ± 95	2585 ± 258	1424 ± 124	1232 ± 263	335 ± 110	124 ± 69
Jennings *et al.* 2012a		0.4-15 km·h^{-1}	>15 km·h^{-1}				
	D	7686^* ± 400	1734^* ± 177				
	M	7363^ ± 290	2554^` ± 134				
	F	7405* ± 472	2189*` ± 456				
	All	7505 ± 416	2117 ± 444				
Jennings *et al.* 2012b		0.4-15 km·h^{-1}	>15 km·h^{-1}				
	I	7441 ± 511	2294 ± 433				
	N	6905 ± 447	1652 ± 416				
Buglione *et al.* 2013		0.1-6 km·h^{-1}	6.1-11 km·h^{-1}	11.1-14 km·h^{-1}	14.1-19 km·h^{-1}	19.1-23 km·h^{-1}	>23 km·h^{-1}
	E	2592 ± 396	2639 ± 339	1557 ± 218	1295 ± 184	344 ± 169	102 ± 74
	S	2555 ± 324	2349 ± 569	1439 ± 389	1286 ± 446	346 ± 134	102 ± 101
Liu *et al.* 2013		<7.6 km·h^{-1}	7.6-11.5 km·h^{-1}	11.5-15.5 km·h^{-1}	15.5-20.5 km·h^{-1}	20.5-29.5 km·h^{-1}	>29.5 km·h^{-1}
	D	2709^ ± 221	1478*^ ± 86	1148*^ ± 139	816*^ ± 134	486*^ ± 77	36* ± 9
	M	2430^ ± 133	1872^ ± 159	1623^ ± 208	1149^ ± 116	608^ ± 89	51` ± 12
	F	2567 ± 138	1774* ± 131	1583* ± 162	1100* ± 135	602* ± 79	72*` ± 16
	All	2580 ± 261	1693 ± 243	1434 ± 321	1013 ± 241	560 ± 126	53 ± 23
Lythe & Kilding 2013		0-11 km·h^{-1}	11.1-19 km·h^{-1}	>19.1 km·h^{-1}			
	F	4796 ± 232	2937 ± 203	701 ± 92			
Vescovi & Frayne 2015		0-8 km·h^{-1}	8.1-16 km·h^{-1}	16.1-20 km·h^{-1}	>20.1 km·h^{-1}		
	D	2958* ± 635	2926 ± 696	551^ ± 188	113 ± 83		
	M	2657 ± 777	3281 ± 839	680^ ± 189	136 ± 72		
	F	2379* ± 637	2858 ± 774	661 ± 142	153 ± 77		
	All	2651 ± 724	3056 ± 797	640 ± 181	135 ± 77		
		19-23 km·h^{-1}	>23 km·h^{-1}				

Authors		Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6
White & MacFarlane 2015a	All	457 ± 57	114 ± 53				
Kim *et al.* 2016		<6 km·h^{-1}	6-12 km·h^{-1}	12-14 km·h^{-1}	18-24 km·h^{-1}	>24 km·h^{-1}	
	All	1562 ± 263	1983 ± 314	580 ± 126	773 ± 166	371 ± 74	
Vescovi 2016		0-8.0 km·h^{-1}	8.1-16.0 km·h^{-1}	16.0-20.0 km·h^{-1}	>20.1 km·h^{-1}		
	D	2342^* ± 542	2287* ± 440	405 ± 81	101 ± 35		
	M	1799`^ ± 542	2339` ± 682	496 ± 163	91 ± 56		
	F	1241`* ± 355	1538`* ± 376	389 ± 142	106 ± 52		
	All	1756 ± 644	2051 ± 639	434 ± 144	98 ± 49		
Ihsan *et al.* 2017		<15 km·h^{-1}	>15 km·h^{-1}				
	All	4406 ± 390	826 ± 151				
Sunderland & Edwards 2017		>15.5 km·h^{-1}	>20 km·h^{-1}				
	D	1364*^ ± 202	421*^ ± 79				
	M	1589^ ± 212	483^ ± 78				
	F	1635* ± 314	513* ± 127				
	All	1529 ± 274	472 ± 104				
Casamichana *et al.* 2018		15.1-18.9 km·h^{-1}	>19 km·h^{-1}	>24 km·h^{-1}	>30 km·h^{-1}		
	D	1040^* ± 309	364^* ± 136	50^* ± 40	0^ ± 1		
	M	1581^ ± 307	645^ ± 167	144^ ± 82	4^ ± 8		
	F	1520* ± 458	639* ± 209	137* ± 66	2 ± 6		
	All	1372 ± 432	544 ± 216	109 ± 78	2 ± 6		
McGuinness *et al.* 2018		0-7.9 km·h^{-1}	8-10.9 km·h^{-1}	11-15.9 km·h^{-1}	>16 km·h^{-1}	>20 km·h^{-1}	
	D	2448 ± 384	969 ± 151	1214 ± 166	412* ± 93	140* ± 62	
	M	2000 ± 155	883 ± 142	1313 ± 173	424` ± 93	121` ± 57	
	F	1823 ± 370	775 ± 149	1287 ± 143	503*` ± 85	154*` ± 57	
	All	2113 ± 369	883 ± 144	1269 ± 185	443 ± 88	137 ± 59	
Polglaze *et al.* 2018		>15.5 km·h^{-1}					
	D	1087^ ± 267					
	M	1283`^ ± 139					
	F	1101` ± 227					

Authors		Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6
	All	1156 ± 235					
Vescovi & Klas 2018		0-8 km·h^{-1}	8.1-16 km·h^{-1}	16.1-20 km·h^{-1}	>20 km·h^{-1}		
	All	3107 ± 592	3083 ± 575	524 ± 161	124 ± 43		
McMahon & Kennedy 2019		0-11.0 km·h^{-1}	>11 km·h^{-1}				
	D	4449* ± 995	733*^ ± 204				
	M	4255' ± 750	1045^` ± 273				
	F	3599*' ± 749	946*` ± 267				
	All	4074 ± 887	937 ± 284				
Morencos *et al.* 2019		>15 km·h^{-1}	>21 km·h^{-1}				
	All	901 ± 252	859 ± 225				

Significance difference ($p < 0.05$) between: *forward and defense, ^midfield and defense, `forward and midfield. Grey highlighting indicates data are by position not per athlete. D: Defense, M: Midfield, F: Forward, I: International, N: National, E: Elite, S: Subelite.

Table 4.6: Internal Training Load

Authors	
Lythe & Kilding 2011	Overall mean HR: 85.3 ± 2.9% (F: 85.6 ± 2.7%, M: 86.5 ± 2.6%, D: 84.7 ± 2.6%). Percent of match <75% max HR: 9.9 ± 5.7% (D: 8.9 ± 5.1%, M: 8.0 ± 3.3%, F: 11.8 ± 7.3%). Percent of match at 75-84% max HR: 29.3 ± 12.2% (D: 29.1 ± 11.5%, M: 29.8 ± 9.7%, F: 27.3 ± 14.6%). Percent of match at 84-95% max HR: 56.4 ± 13.0% (D:60.2 ± 13.9%, M: 57.2 ± 9.8%, F: 53.1 ± 13.9%). Percent of match >95% max HR: 4.3 ± 6.6% (D: 1.7* ± 3.2%, M:5.1 ± 5.9%, F:7.8*±9.0%).
Buglione *et al.* 2013	Overall mean HR: Elite - 84.5 ± 3.7%, Sub-elite - 85.8 ± 2.8%. Blood lactate concentration (elite athletes): 1st half - 4.3 ± 1.7 $mmol \cdot L^{-1}$, 2nd half- 5.3 ± 2.7 $mmol \cdot L^{-1}$.
Lythe & Kilding 2013	Forwards only (substitution pattern- 5 players for 3 positions). Overall mean HR 85.4±6.7%. Percent of match <75% max HR: 10.5±6.6%. Percent of match 75-84% max HR: 28.6 ± 18.2%. Percent of match at 85-95% max HR: 56.4±17.2%. Percent of match >95% max HR: 4.7±7.0%.
Vescovi *et al.* 2016	Overall mean HR: 89 ± 4% (D: 90 ± 4%, M: 88 ± 5 %, F: 90 ± 3%). Time <80% max HR (minutes): 5 ± 3 (D: 5 ± 4, M: 5 ± 4, F: 4 ± 2). Time 80-90% max HR (min): 11 ± 9 (D: 16 ± 10, M: 11 ± 9, F: 6 ± 5). Time >90% max HR (min): 23 ± 11 (D: 31* ± 13, M: 21 ± 10, F: 19* ± 7).
Crewther *et al.* 2017	sRPE: 343 ± 149 AU.
Perrotta *et al.* 2017	sRPE: 558±67 AU. Training impulse (Banister with recovery estimation model): 235±56 AU. Including warmup.
McGuinness *et al.* 2018	Time <69% max HR (minutes): 9 ± 5 (D: 7* ± 5, M: 9 ± 5, F: 9* ±5). Time 70-84% max HR (min): 24 ± 7 (D: 21* ± 7, M: 24\` ±7, F: 28*\` ± 7). Time >85% max HR (min): 14 ± 6 (D: 14 ± 6, M: 12 ± 4, F: 15 ± 5). Time >90% max HR (min): 18 ± 10 (D:24* ± 1, M: 18 ± 10, F:12* ± 11).
Perrotta & Warburton 2018	Overall mean HR: 79.0 ± 5.6%. Training load (Polar): 662± 148 AU. Training load (Edwards): 713±137 AU. sRPE: 2280 ± 296.
Vescovi & Klas 2018	Time <80% max HR (minutes): 56 ± 10. Time at 80-90% max HR (min): 21 ± 3. Time >90% max HR: 15 ± 10.
Krueger *et al.* 2019	RPE: 6.2 ± 1.5 AU.
Vescovi *et al.* 2019	Training impulse (Stagno): 312± 61 AU. sRPE: 769±275 AU. Including warmup.

Significance difference ($p < 0.05$) between: *forward and defense, ^midfield and defense, \`forward and midfield.
HR: Heart rate, sRPE: Session rating of perceived exertion. AU Arbitrary units. F: Forward, M: Midfield, D: Defense.

4.3.1 Study characteristics

A summary of study and participant characteristics is provided in Table 4.1. Of the 31 studies identified for inclusion, there was an equal divide across sexes, with 15 studies on female athletes and 16 on male athletes. Likely due to the increased monitoring of these athletes, there was a disproportionate number of studies on elite athletes, particularly international and junior international athletes, with 24 of 31 studies providing data on these groups. As a result, metanalyses were performed only for elite athletes. Similarly, as the format of most international hockey competition during this period was tournaments and test series, the majority of data was collected in these settings, with only seven studies providing data exclusively on league play. In terms of the training load data collected, most studies measured external load metrics, with all but four studies using some form of Global Positioning System (GPS) device. In contrast, only ten studies considered internal training load in the form of heart rate data, and six studies measured internal training load using session rating of perceived exertion (sRPE).

4.3.2 Risk of bias

The mean risk of bias score was 8.6 ± 0.9, with a range of 6-10 out of a maximum possible score of 11 which would have indicated minimal risk of bias (Table 4.2). The questions on the quality assessment checklist with the fewest positive responses were items 4 and 11, regarding inclusion/exclusion criteria and collecting data on more than one team. Specifically, only eight studies incorporated athletes from at least two teams and nine studies included information on the health and injury status of participants as part of their inclusion/exclusion criteria. However, particularly in elite hockey populations, one would expect that athletes would only be participating in competition if they were deemed healthy by the team doctor or physiotherapist. No studies were excluded from the review due to risk of bias.

4.3.3 Results of individual studies

A summary of the training load measures reported in each study is shown in Table 4.3. Although this list is not exhaustive of all training load measures described across studies, it incorporates the most commonly used methods of measuring training load in hockey populations. Monitored in 25 studies, total distance was the most frequently reported measure of training load, and 24 studies measured distances in various speed categories. Less frequently reported measures of external load were playerload (n = 5) and metabolic power (n = 3). Although most studies measured external training load (n = 28), only twelve studies incorporated measures of internal load. When internal load was reported, it was most often measured in terms of heart rate (n = 8), either as an average value or time within various thresholds. Although the easiest to measure, the subjective rating of perceived exertion (RPE) was reported in only six studies. The majority of studies (n = 20) performed analysis by playing position. The most common positional categories considered were forward/striker, midfield, and defense; however, some studies provided further distinctions by separating attacking and defensive midfielders or halfbacks and fullbacks. Additionally, 11 studies presented data separately across either halves or quarters of competition.

A summary of external training load results is provided in Table 4.4. The external load measures of total distance, workrate, and maximal speed were selected as summary measures due to their frequent reporting and consistent definitions across studies. In contrast, what constituted a sprint or an acceleration varied notably across studies, making direct comparisons of these results inappropriate. Even for the straightforwardly defined measures of total distance and workrate, there were slight discrepancies in reporting with some studies reporting total distance covered by position rather than by individual athletes and others incorporating time spent on the bench in workrate calculations.

Table 4.5 provides a breakdown of the distances covered by athletes in various speed zones, as well as the definitions for those speed zones. Despite previous recommendations for the adoption of uniform speed zone thresholds (Dwyer and Gabbett, 2012), there has been no consistent approach to speed zone definitions across studies. Studies utilized up to six speed zones, with an average of 3.5 ± 1.6 zones reported.

A summary of internal training load results is provided in Table 4.6. Due to the varied measurement methods and means of reporting, these results are provided in narrative rather than tabular form. Even when the same internal training load measures were used, the reporting of results varied, with some reporting time spent in heart rate zones while others provided an average percentage of maximum heart rate. Compared with external load measures, positional breakdowns were less frequently provided for internal load data, with only three studies reporting results separately by position.

Table 4.7: External Training Load in Elite Hockey Metanalysis

	Female	Male
Total Distance (m)	5029 ± 424* (4657 – 5401)	6027 ± 536* (5677 – 6377)
Defense	5167 ± 246 (4889 – 5445)	6346 ± 615 (5650 – 7042)
Midfield	4817 ± 248 (4536 – 5099)	6190 ± 650 (5455 – 6925)
Forward	4115 ± 802 (3208 – 5023)	5437 ± 428 (4953 – 5921)
Workrate ($m \cdot min^{-1}$)	115 ± 8 (107 – 122)	125 ± 8 (119 – 132)
Defense	105 ± 6 (99 – 110)	120 ± 5 (114 – 126)
Midfield	118 ± 6 (112 – 124)	134 ± 8 (125 – 142)
Forward	120 ± 12 (108 – 131)	131 ± 18 (111 – 151)

Mean ± SD. (95% confidence interval).

*Significant difference in total distance between male and female hockey ($p = 0.003$)

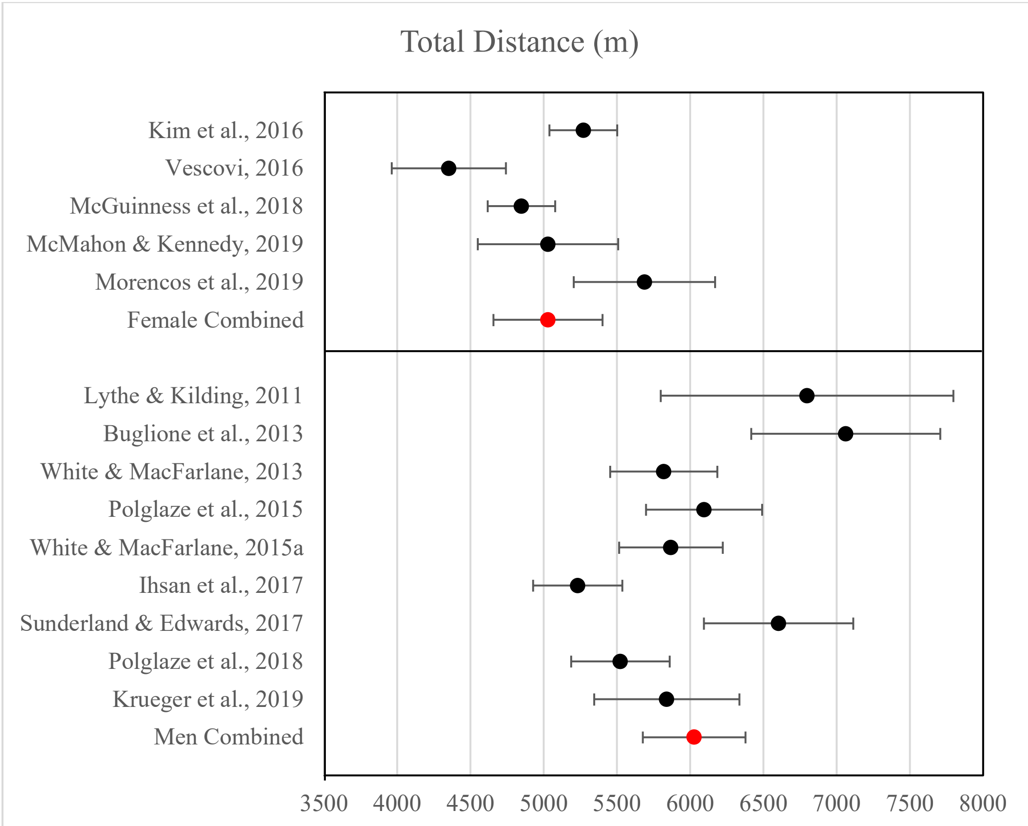


Figure 4.3: Total Distance in Elite Hockey

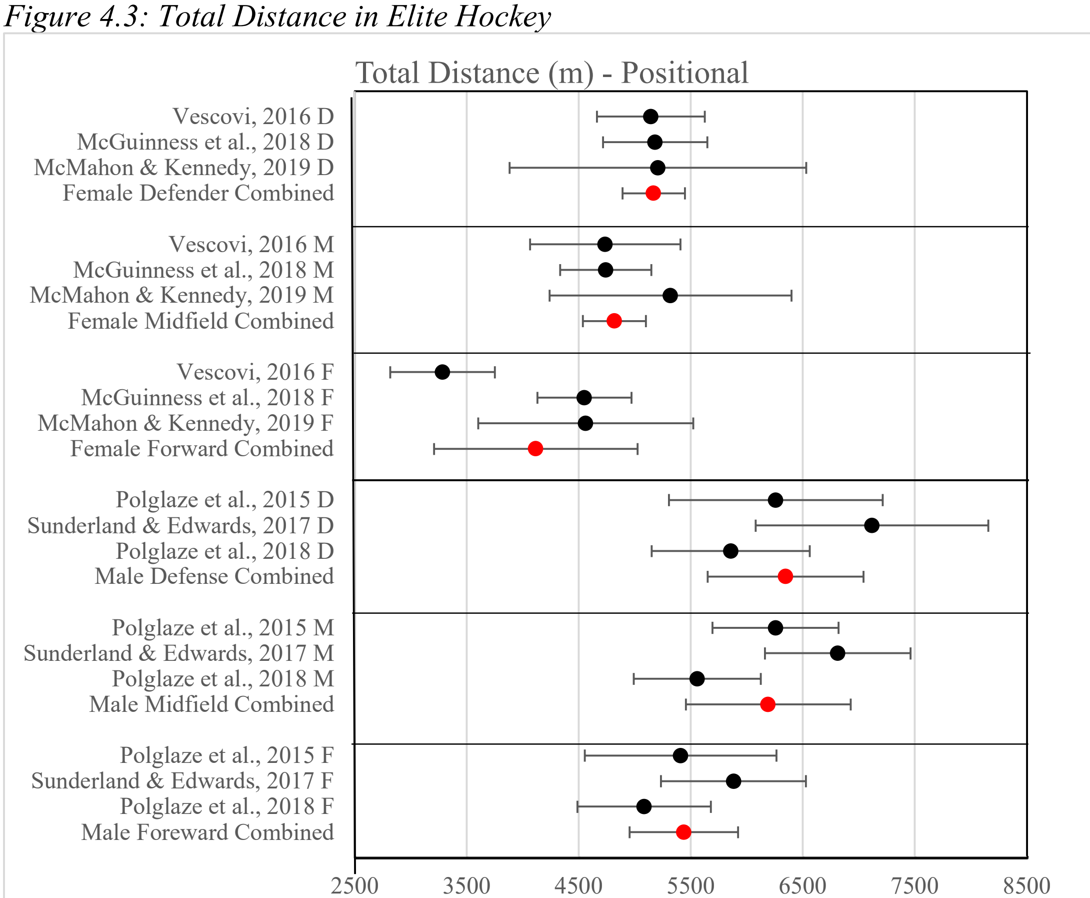


Figure 4.4 Total Distance in Elite Hockey by Position

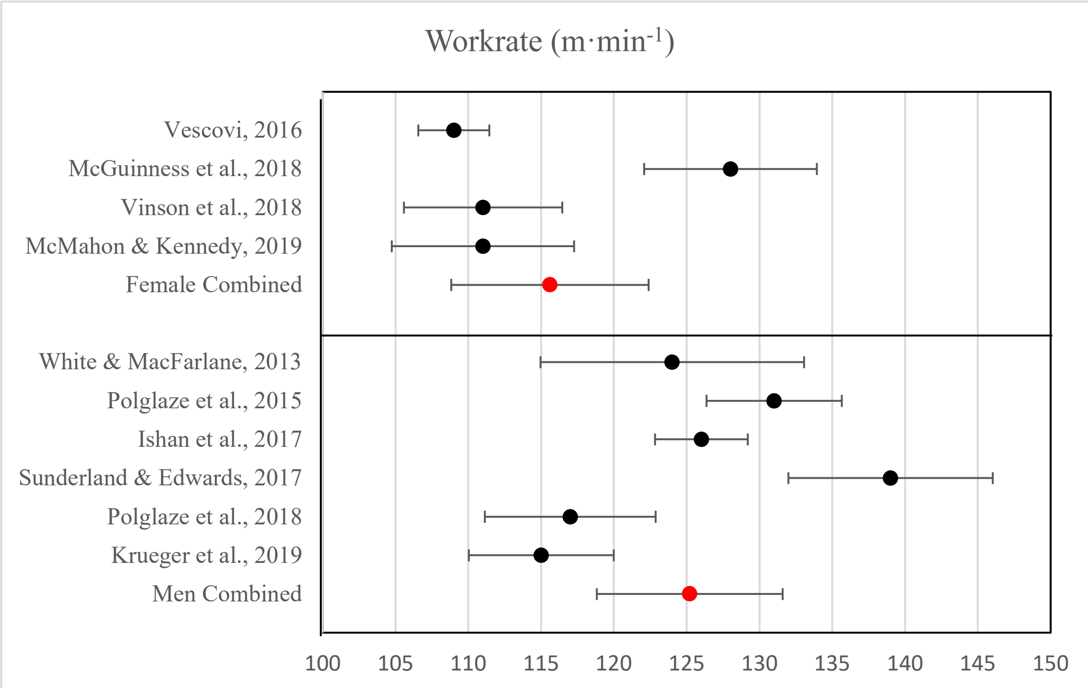


Figure 4.5: Workrate in Elite Hockey

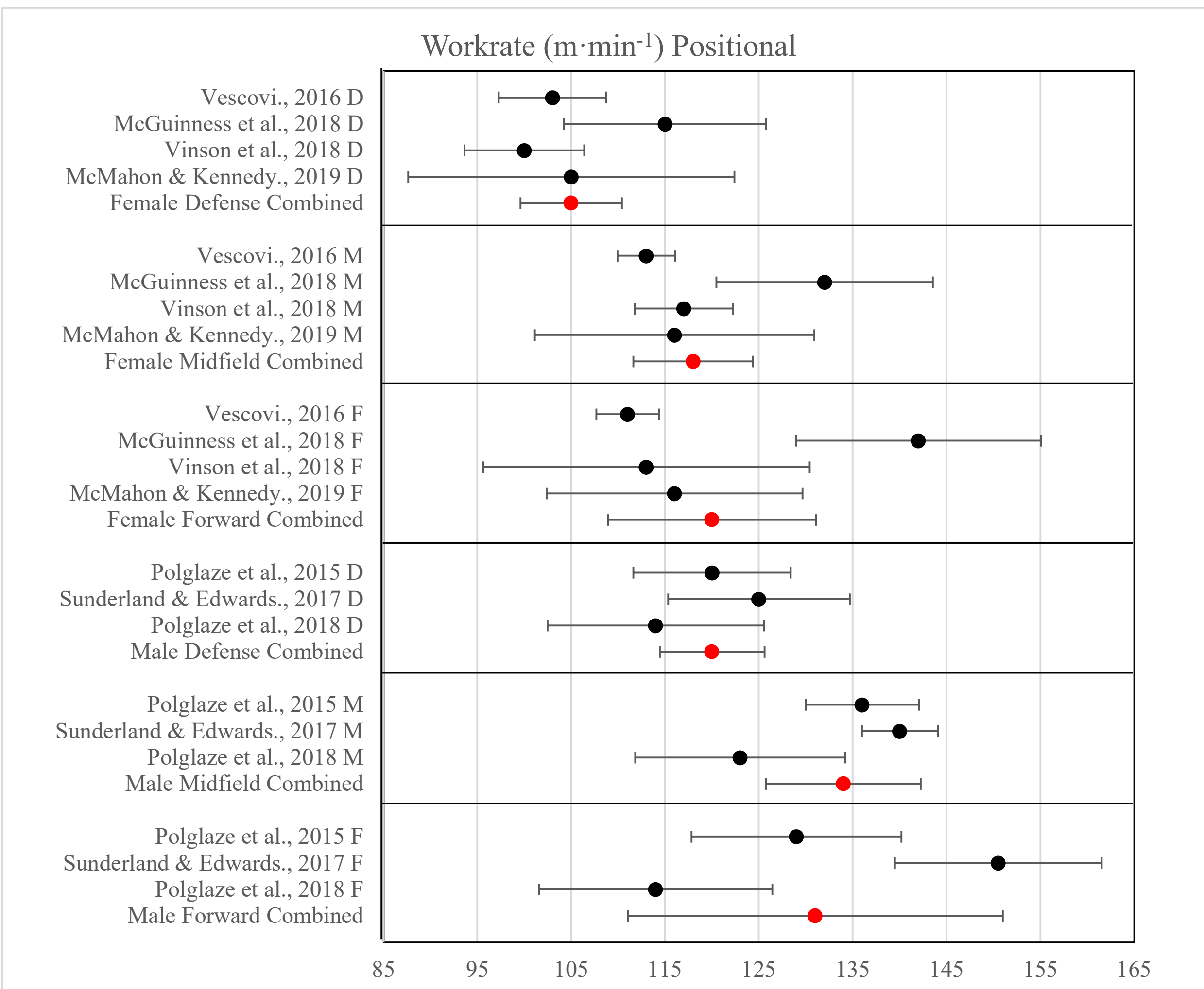


Figure 4.6: Workrate in Elite Hockey by Position

4.3.4 Results of synthesis

Results of the meta-analyses for total distance and workrate in elite hockey are summarized in Table 4.7. The studies included in each meta-analysis are shown in the corresponding forest plots (Figures 4.3-4.6). The forest plots clearly demonstrate the range and mean values both for total distance and workrate across elite hockey populations. As a result of the large between-studies variance, which was incorporated as part of the random effects model, the confidence intervals derived through the meta-analyses were, in some cases, wider than those of the individual studies. However, as many of the studies incorporated data from only one team, this increased variance around the mean is to be expected as external load values will likely vary more for elite hockey populations rather than within individual teams.

Figures 4.3 and 4.4 illustrate the total distance covered in male and female elite hockey. Male athletes covered significantly more total distance than female athletes in hockey competition (M: 6027 ± 536 m, F: 5029 ± 424 m, $p = 0.003$), with this difference likely due to the physiological differences between male and female athletes. Additionally, despite many individual studies reporting significant differences in distances covered between positions, one-way ANOVA of the meta-analysis results found no significant positional differences for male athletes ($p = 0.196$) or female athletes ($p = 0.104$). However, the small sample size likely contributed to this lack of significance, with few studies reporting total distance by position. Even though the differences were not significant, the results demonstrate a similar trend in both male and female hockey with defenders covering the greatest total distance, followed by midfielders and then forwards.

Results on workrate in male and female elite hockey are included in figures 4.5 and 4.6. Unlike total distance, there was a trending but not significant difference in workrate between male and female hockey (M: 125 ± 8 $m \cdot min^{-1}$, F: 115 ± 8 $m \cdot min^{-1}$, $p = 0.079$). However, again, this lack of significance may have been due to a decreased sample size, with fewer studies reporting workrate than total distance. Positional breakdowns revealed a trending but not significant difference in workrate across playing positions in female hockey ($p = 0.057$) with the greatest differences between defenders and forwards ($p = 0.072$) and defenders and midfielders ($p = 0.105$). No significant differences were found across positions in male hockey ($p = 0.355$). Unlike total distance for which defenders had the highest loads, workrate was similar for midfielders and forwards and notably lower for defenders, in both male and female athletes.

4.4 Discussion

This review aimed to synthesize the evidence and draw conclusions on internal and external training load in male and female competitive hockey players. In doing so, the main findings were as follows. Elite male athletes covered significantly more distance in matches than elite female athletes (M: 6027 ± 536 m, F: 5029 ± 424 m, $p = 0.003$), with workrate also notably higher in elite male athletes (M: 125 ± 8 $m \cdot min^{-1}$, F: 115 ± 8 $m \cdot min^{-1}$, $p = 0.079$). In male and female hockey, defenders covered greater distances and had lower workrates than midfielders and forwards; however, these differences were not significant. Other training load measures were less frequently reported, with definitions varying across studies. Standardized definitions and thresholds for training load measures need to be adopted to allow for integrative research and further comparison across hockey populations.

4.4.1 External training load

External training load was the most common method of assessing athlete load during competition with 28 studies reporting some form of external load data. External load was primarily measured via GPS or GNSS units, with only one study using time-motion analysis (Liu *et al.*, 2013). Most studies used Catapult Sports systems for external load monitoring, with a singular study (McGuinness *et al.*, 2019) using GPS monitors not produced by either Catapult Sports or GPSports, which has since been acquired by Catapult Sports. In addition, the majority of studies were performed with 5 Hz or 10 Hz GPS units, which have been repeatedly shown to be more accurate than 1 Hz models (1 Hz: 4 studies, 5 Hz: 13 studies, 10 Hz: 9 studies, 15 Hz: 1 study) (Scott, Scott and Kelly, 2016).

4.4.1.1 Total distance

A straightforwardly defined and relatively easy to monitor measure of external training load, total distance was recorded in 25 studies. Mean total distance differed notably across studies, with the highest average distance reported for individuals (7675 m) (Morencos *et al.*, 2018) over 175% of the lowest reported distance (4351 m) (Vescovi, 2016). The variance of total distance also markedly differed across studies, with standard deviations ranging from 479 m (Ihsan *et al.*, 2017) to 2009 m (Lythe and Kilding, 2011), indicating differences in the spread of distance values across hockey populations. Although also impacted by sample size and homogeneity of the population, this large range of standard deviations would suggest that athletes in some teams or cohorts cover similar distances during matches, while in other groups there is greater disparity in the work performed by various athletes. This conclusion is further reinforced by the positional breakdowns, with significant differences in total distance across groups in eight out of the ten studies reporting positional breakdowns with sample sizes and standard deviations.

Although individual studies demonstrated significant positional differences in total distance, there were no significant positional differences found in the meta-analyses of male ($p = 0.196$) or female hockey ($p = 0.104$). Although these results may seem contradictory, there was little consensus across studies on which positions covered the greatest total distance. Additionally, few studies reported total distance by position, with smaller sample sizes resulting in wider confidence intervals. However, a trend did emerge in both male and female elite hockey with defenders covering the greatest total distance followed by midfielders and forwards. This positional difference in total distance is often attributed to the fact that defenders have been repeatedly shown to play the most minutes with forwards playing the least (Polglaze *et al.*, 2015; Vescovi and Frayne, 2015; Vescovi, 2016; Casamichana *et al.*, 2018; McGuinness *et al.*, 2019; McMahon and Kennedy, 2019). Although these general trends have emerged in elite hockey, the few studies reporting positional distances in sub-elite hockey show different distributions. For example, Liu *et al*. reported mean total distance, measured by time-motion analysis in male national level Chinese games athletes to be significantly higher ($p < 0.05$) for forwards (7709 m) and midfielders (7733 m) than defenders (6771 m) (Liu *et al.*, 2013). Similarly, Morencos *et al.* also reported a greater total distance in midfielders (7867 m) than defenders (7533 m) in sub-elite male hockey, although this result was not significant (Morencos *et al.*, 2018). Higher total distance in midfielders rather than defenders in sub-elite hockey may be due to a greater discrepancy in skill and fitness at this level, with the better athletes playing the most minutes and being positioned in the middle of the field where they can have the greatest impact on play. However, more research will be needed to confirm this theory. Finally, it is important to note that despite these trends emerging in distance distributions across positions in elite and sub-elite hockey, there will always be exceptions due to individual athletes, and the tactics and substitution strategies of specific teams.

Comparing male and female athletes, elite male athletes on average cover an additional kilometer per match compared to elite female athletes (M: 6027 ± 536 m, F: 5029 ± 424 m, $p = 0.003$). However, it is important to note that unlike some other intermittent ball sports, such as lacrosse and basketball, there are no differences in the rules, equipment, or match time in male and female hockey. Although, the results of the metanalysis clearly illustrate a significant difference on average between male and female hockey, not all individual studies followed this trend. For example, the average total distance measured in members of the Spanish Women's National Team (5687 ± 905 m) (Morencos *et al.*, 2019) was greater than the average measured in the Singapore Men's National Team (5232 ± 479 m) (Ihsan *et al.*, 2017) and the Australian Men's National Team (5523 ± 632 m) (Polglaze *et al.*, 2018). However, the results of the meta-analysis clearly demonstrated that for elite hockey on the whole, male athletes cover greater total distances than female athletes.

In addition to reporting total distance, several studies examined the relationship between distance covered and other variables. For example, Liu *et al.* reported that total distance was significantly lower in the second half a match compared to the first half in sub-elite male hockey (Liu *et al.*, 2013). However, this finding was in contrast to the results of another study also on sub-elite male hockey athletes, where total distance was found to be similar across quarters (Morencos *et al.*, 2018). Therefore, the maintenance of total distance across halves or quarters appears to vary across hockey populations, perhaps due to the fitness level of the athletes. The relationship between total distance covered and opponent quality was also considered, with total distance shown to be significantly lower when there was a mismatch in team rankings (White and Macfarlane, 2015b) or when a team won by a large margin (Jennings *et al.*, 2012b). Finally, in a study on junior international athletes, the distance covered by athletes during warm-up prior to the game was shown to be substantial, making up 27% of the overall distance covered on a matchday.

4.4.1.2 Workrate

Given that rolling substitutions in hockey result in individual athletes playing for varied minutes, workrate provides a relative measure to consider distance covered in hockey. Workrate is most frequently measured in meters per minute and is a measure of average speed, generally based on the time that an athlete spends on the pitch. As one of the more frequently monitored and most consistently defined measures of external training load, workrate provides a good metric for comparison across studies. Although not significant, there was a notable difference ($p = 0.079$) in workrate between elite male and female hockey, with male athletes averaging an additional 10 $m \cdot min^{-1}$ than female athletes. (M: 125 ± 8 $m \cdot min^{-1}$, F: 115 ± 8 $m \cdot min^{-1}$). Fewer studies reported workrate than total distance; therefore, the lack of significance of this finding is likely related to the smaller sample size. However, the finding that male athletes average higher speeds than female athletes corresponds with the result that elite male athletes cover greater total distances during competition.

In terms of workrate across positions, the results of the meta-analysis suggest that workrate in elite hockey is similar in midfielders and forwards, but lower in defenders. Specifically, in elite female hockey, mean workrates in forwards (120 ± 12 $m \cdot min^{-1}$) and midfielders (118 ± 6 $m \cdot min^{-1}$) were notably higher than in defenders (105 ± 6 $m \cdot min^{-1}$) (F v D: $p = 0.072$, M v D: $p = 0.105$). Similarly, in elite male hockey, workrate was similar in forwards (134 ± 7.5 $m \cdot min^{-1}$) and midfielders (131 ± 18 $m \cdot min^{-1}$) but much lower in defenders (120 ± 5 $m \cdot min^{-1}$). However, these differences were not significant ($p = 0.355$), as a result of the small sample size leading to large standard deviation values. Despite this, there is good evidence of positional differences based on workrate across hockey populations, with all studies reporting workrate across positions ($n = 9$) having at least one significant difference between positions. The mean workrate for defenders was the lowest in all studies, with 8 studies reporting significant differences ($p < 0.05$) between midfielders and defenders (Polglaze *et al.*, 2015; Vescovi and Frayne, 2015; Vescovi, 2016; Sunderland and Edwards, 2017; Morencos *et al.*, 2018; Polglaze *et al.*, 2018; Vinson, Gerrett and James, 2018; McGuinness *et al.*, 2019). Additionally, there were significant differences ($p < 0.05$) between forwards and defenders in six studies (Polglaze *et al.*, 2015; Vescovi and Frayne, 2015; Vescovi, 2016; Sunderland and Edwards, 2017; McGuinness *et al.*, 2019). There is less evidence on positional differences in workrate between midfielders and forwards with the results often very similar between these groups and only four studies indicating significant differences ($p < 0.05$) (Polglaze *et al.*, 2015; Sunderland and Edwards, 2017; McGuinness *et al.*, 2019; Polglaze *et al.*, 2018). However, in all cases where significant differences were found, workrate was higher in midfielders than in forwards (Polglaze *et al.*, 2015; Sunderland and Edwards, 2017; McGuinness *et al.*, 2019;

Polglaze *et al.*, 2018). In conclusion, the evidence would suggest that workrate tends to be similar between midfielders and forwards and lower in defenders.

Workrate has also been considered in relation to various other study outcomes including rule changes, competition level, and analysis technique. Between the 2014 and 2015 international hockey seasons, changes were made to the format of hockey, with four 15-minute quarters, replacing two 35-minute halves, and additional in-game clock stoppages occurring prior to penalty corners and following goals (McMahon and Kennedy, 2019). As a result of these differences, workrate was shown to substantially increase by 7.5% (effect size- ES: 0.57), despite the amount of time that players spent on the pitch only trivially decreasing by 2.1% (ES: -0.08) (McMahon and Kennedy, 2019). Therefore, the results of this study would suggest that athlete workrate increased following the 2015 rule changes. However, as this study was only performed on one team, it is important to consider that other changes such as fitness differences or tactical strategies may have contributed to the difference in workrate between years. Within elite hockey, workrate has also been shown to differ significantly between international and top national-level competition (Jennings *et al.*, 2012b). Specifically, there were moderate to very large differences in workrate for each position between male athletes representing Australia in the international Champions Trophy and male athletes competing in the Australian Hockey League (Forward: 8%, ES-1.44; Midfield: 11.1%, ES-1.93; Defender: 11.9%, ES-1.87) (Jennings *et al.*, 2012b). Finally, as would be expected due to rolling substitutions in hockey, workrate was shown to be significantly higher when calculated using a time-on-pitch rather than a full game analysis procedure ($p < 0.001$) despite the distance covered remaining almost the same (White and MacFarlane, 2013). Therefore, in order to provide an accurate indicator of athlete speed during competition, it is critical to calculate workrate using only the time that each athlete spends on the pitch rather than the full game, as was done in two studies (White and Macfarlane, 2015a; Morencos *et al.*, 2019)

4.4.1.3 Distance in speed zones

Second to total distance, distance in speed zones was frequently reported across studies with 77% of studies reporting data on this training load measure. However, despite the frequent measurement of distance in speed zones, there is little to no consensus on cutoff values for speed zones. Consequently, without standardized thresholds, it is almost impossible to compare results across studies and develop an overall understanding of the distance hockey athletes typically cover at various speeds. Similarly, while some studies reported the number of sprints and the distance covered at a sprint, there is no consensus on what speed is considered to be a sprint and what the requirements are for an individual burst to be counted as a sprint repetition. Therefore, in order to allow for comparisons across future research, it is recommended that researchers work together to develop a consensus on standardized speed zones for use in hockey and adopt these thresholds across studies.

Although the varying speed zone thresholds prevent direct comparison, it is possible to draw some conclusions from the available data. Specifically, athletes covered most of their total distance at lower speeds (below 15 $km \cdot h^{-1}$) with at least 2.5km covered at a walk (below approximately 6-8 $km \cdot h^{-1}$). Furthermore, distance covered in speed zones generally tends to decrease as speed increases, with athletes covering less distance at what could be considered sprint speeds (>20 $km \cdot h^{-1}$) compared to fast running (approximately 15-19 $km \cdot h^{-1}$). In terms of positional differences, forwards and midfielders were found to cover a significantly greater distance than defenders in moderate and high speed zones (above approximately15 $km \cdot h^{-1}$), with defenders covering significantly more distance in lower speed zones (Jennings *et al.*, 2012b; Vescovi, 2016; Sunderland and Edwards, 2017; Polglaze *et al.*, 2018; McGuinness *et al.*, 2019; Casamichana *et al.*, 2018). These conclusions on speed zones align with positional results on workrate, with midfielders and forwards averaging higher speeds than defenders while on the pitch.

Distance covered in speed zones has also been analyzed across quarters and over the course of a tournament. In terms of high-intensity distance across quarters, athletes tend to cover greater high-intensity distance in the first and final quarters than in quarters two and three. Specifically, an analysis of Spanish international female hockey athletes found high-intensity distance (>15 $km \cdot h^{-1}$) to be higher in quarter four compared to the second and third quarter (ES: 0.6-1.25) but not compared to quarter 1 (ES: 0.1-0.4) (Morencos *et al.*, 2019). Similarly, a study of 27 female international athletes found that high-intensity distance (>16 $km \cdot h^{-1}$) was significantly lower in the second quarter compared to the first and fourth quarters (McGuinness *et al.*, 2019). These results suggest that despite fatigue, elite female hockey athletes are able to increase their high-intensity external workload at the end of a match. However, these results are contrary to those found in sixteen sub-elite male athletes, where high-speed running distance

(>19 $km \cdot h^{-1}$) was found to be similar across quarters (Morencos *et al.*, 2018). It is possible that this difference may have been due to the sub-elite level of the athletes studied, with lower fitness levels potentially resulting in athletes being unable to increase their output in the final quarter; however, more research will be needed to confirm these findings. Considering distance in speed zones measured during a tournament, fifteen athletes from the Australian men's national team were studied over the course of an international competition consisting of six matches played in nine days (Champions Trophy) (Jennings *et al.*, 2012b). Although between-game variation did occur in high speed running (>15 $km \cdot h^{-1}$) for each position, specifically with higher values in the first match, there was no evident reduction in high speed running over the course of the tournament (Jennings *et al.*, 2012b). These results suggest that despite the very intense and compact schedules of international hockey tournaments, trained athletes are generally able to maintain their high-speed running output throughout.

4.4.1.4 Maximal sprint speed

Nine studies reported the maximal sprint speed of hockey athletes during competition and demonstrated that maximal sprint speed was consistent for both sexes, with female athletes averaging a maximal sprint speed of approximately 24 $km \cdot h^{-1}$ (Vescovi, 2014; Vescovi and Frayne, 2015) and male athletes approximately 27 $km \cdot h^{-1}$ (Lythe and Kilding, 2011; White and MacFarlane, 2013; Kim, Cha and Park, 2018). Compared to other measures of external training load, maximal sprint speed did not vary largely by position with two studies not reporting any significant positional differences (Vescovi, 2014; Vescovi and Frayne, 2015). These results would suggest that although athletes in different positions tend to have different average speeds, or workrates, the maximal speed achieved is relatively consistent. However, given that both of these studies were performed on female athletes under the age of 23, more research is needed in other hockey populations.

Maximal sprint speed has been used to classify athletes and to determine relative speed thresholds. A study of sixteen male hockey athletes competing at the national level found that the mean distances covered at moderate-speed running, high-speed running, and very high-speed running were all significantly different ($p < 0.001$) when generic versus individual thresholds were applied (based on maximal running speed measured either during hockey or non-hockey conditioning) (Casamichana *et al.*, 2018). Furthermore, when individual thresholds were used, distances covered at very high speeds were significantly underestimated in defenders ($p < 0.001$) and overestimated in midfielders and forwards ($p < 0.001$), suggesting that individualizing speed thresholds may impact comparisons between positions (Casamichana *et al.*, 2018). In addition, a study of junior international athletes found that athletes only achieved approximately 90% of their maximal sprint speed, as measured during a 35 m sprint assessment, during hockey competition (Vescovi, 2014). Athletes were also classified into two groups (slower and faster) based on their maximum sprint speed, with slower athletes reaching similar mean sprint speeds to the faster athletes during competition by running at a higher percentage of their overall maximum speed.

4.4.1.5 Accelerations and decelerations

In addition to considering distance and speed, accelerations and decelerations provide further information on the external loads of hockey athletes. However, as with speed zones, comparisons of accelerations and decelerations between studies are made difficult due to the varied definitions and zones implemented across studies. For example, Chesher *et al.* used receiver operator characteristic curves to determine deceleration intensity bands of low intensity (-3 to -5.99 $m \cdot s^{-2}$), medium intensity (-6 to -8.99 $m \cdot s^{-2}$), high intensity (-9 to -11.99 $m \cdot s^{-2}$), and very high intensity (< -12 $m \cdot s^{-2}$), whereas Morencos *et al.* defined decelerations as low intensity (-1 to -1.9 $m \cdot s^{-2}$), moderate intensity (-2 to 2.9 $m \cdot s^{-2}$), and high intensity (<-3 $m \cdot s^{-2}$) (Morencos *et al.*, 2018; Chesher *et al.*, 2019; Morencos *et al.*, 2019). Despite these differences, there were noteworthy findings on accelerations and decelerations in individual studies. For example, moderate (2 - 2.9 $m \cdot s^{-2}$) and high intensity (> 3 $m \cdot s^{-2}$) accelerations and high intensity decelerations (< -3 $m \cdot s^{-2}$) were found to significantly decrease ($p < 0.05$) in quarter 4 comparted to quarter 1 in male hockey athletes competing at a national level, while low intensity accelerations (1 - 1.9 $m \cdot s^{-2}$) significantly increased ($p < 0.05$) and total distance and high speed running remained approximately the same (Morencos *et al.*, 2018). Similarly, a study of fifteen male international athletes found that decelerations were significantly more intense in the first half of matches than the second ($p < 0.001$) (Chesher *et al.*, 2019). Therefore, it has been suggested that acceleration and deceleration may be highly sensitive indicators of fatigue in hockey (Morencos *et al.*, 2018). Finally, despite many studies reporting clear positional differences in total distance, workrate, and distance in speed zones, no significant positional differences were found in the number or intensity of decelerations between positions in elite male hockey (Chesher *et al.*, 2019).

4.4.1.6 Other measures of external training load

Several other measures of external training load were also reported, including measures of metabolic power and playerload. Metabolic power summarizes the demands of intermittent activity by incorporating the energy cost of accelerations to approximate instantaneous metabolic power and overall energy expenditure despite athletes' frequent changes of speed (Polglaze *et al.*, 2018). Mean power ($W \cdot kg^{-1}$) and distances covered by female junior international athletes in high (20 - 35 $W \cdot kg^{-1}$), elevated (35 - 55 $W \cdot kg^{-1}$), and maximal (> 55 $W \cdot kg^{-1}$) metabolic power categories were found to be significantly higher in U21 athletes compared to U17 athletes Vescovi, 2016). Additionally, distance covered at maximal metabolic power largely correlated ($r = 0.624$, $p < 0.001$) with athletes' scores on the Yo-Yo intermittent recovery fitness test (Vescovi, 2016). In elite male hockey, more than 45% of all athlete energy expenditure was shown to be at high intensities (> 20 $W \cdot kg^{-1}$), with no significant difference ($p > 0.05$) in energetic variables found between halves or between games over the course of a tournament (Polglaze *et al.*, 2018). Playerload is an accelerometer-based measure, most frequently defined as the square root of the sum of the squared instantaneous rates of change in all three planes, divided by 100 (Polglaze *et al.*, 2015). However, some studies have used the term Playerload to refer to a propriety software-derived measure aimed at representing both internal and external training load, so it is important to note how Playerload is defined in individual studies (White and Macfarlane, 2015a). In terms of the accelerometer-based definition, Playerload has been shown to be primarily accumulated through running and walking, suggesting that this metric provides little additional information compared to distance-based measures (Polglaze *et al.*, 2015).

Perhaps the most varied in definition across studies, the number of 'sprints' an athlete completes is another commonly reported measure of external training load. Although related, this measure is different from maximal sprint speed or distance covered in speed zones because it provides a count rather than a speed or distance. However, as is the case with other training load measures, what constitutes a sprint, in terms of speed, duration, and consecutive efforts differs vastly between studies and, consequently, will not be examined further here. Other external training load measures that were reported include work to rest ratios, equivalent distance, equivalent distance index, and repeat sprints, indicating the wide range of training load measures that have been used to summarize the demands of hockey competition.

4.4.2 Internal training load

Internal training load is any measure of the physiological demands of exercise and is commonly monitored via heart rate based measures or ratings of perceived exertion (RPE). The simplest summaries of heart rate data are average percentage of maximum heart rate and time in heart rate zones. Average percentage of maximum heart rate (%max HR) was found to be 85 ± 3% (Lythe and Kilding, 2011), 85 ± 4% (Buglione *et al.*, 2013) and 85 ± 7% (forwards only) (Lythe and Kilding, 2013) in three studies of male international athletes, showing very strong consensus on average heart rate values in this population. However, in female athletes there have been more varied values reported with an average of 79 ± 6% in female international athletes (Perrotta and Warburton, 2018) and 89 ± 4% in female junior international athletes (Vescovi, Klas and Mandic, 2019). In terms of positional differences, two studies have reported that defenders spent significantly more time ($p < 0.05$) above 90% max HR than forwards in female international (McGuinness *et al.*, 2019) and junior international hockey (Vescovi, 2016). However, defenders have also been shown to spend significantly less time ($p < 0.05$) than forwards above 95% max HR (Lythe and Kilding, 2011). When considered collectively, these results suggest that defenders spend long periods of time between 90-95% max HR, whereas forwards are spending the majority of their time above 90% max HR also above the 95% threshold. This result would be consistent with previous findings that forwards cover greater distances at top speeds, presumably leading to higher heart rate values, whereas defenders play for more minutes and cover greater distances at moderate speeds.

An advancement on average heart rate or time in heart rate zones, TRIMP uses an algorithm to summarize heart rate data from across a session into a summary score, weighted based on intensity. Different TRIMP algorithms have been applied in various studies, with results expressed in distinct arbitrary units, making comparisons between studies meaningless (Perrotta, Held and Warburton, 2017; Perrotta and Warburton, 2018; Vescovi, Klas and Mandic, 2019). However, comparisons within studies have found TRIMP to differ across positions in female junior international athletes, with defenders having significantly ($p < 0.05$) higher loads than both midfielders and forwards (Vescovi, 2016). In addition, individual studies have examined the relationship between TRIMP and the subjective session rating of perceived exertion (sRPE) method of measuring internal training load. Specifically, TRIMP calculated via Edward's model and a proprietary Polar Team[2] (Polar Electro, Oy, Kempele, Finland) model found moderately correlated (Edwards: $r = 0.60$, Polar: $r = 0.63$) with sRPE, calculated using a 10-point Borg scale (Perrotta and Warburton, 2018). Notably, a slightly weaker correlation was found between sRPE and time spent above lactate threshold ($r = 0.46$), suggesting that Edward's and Polar's TRIMP may be better measures of internal load in hockey than time above lactate threshold (Perrotta and Warburton, 2018).

Several studies have also examined the relationship between sRPE and wellness, and the response of sRPE to other variables. A study of female junior international athletes over the course of a 16-day training and competition tour found no clear relationship between TRIMP or sRPE and wellness scores assessed using the Total Quality Recovery Scale (Vescovi, Klas and Mandic, 2019). However, a different study found pre-game wellness scores to be very strongly correlated with total distance adjusted based on sRPE ($r = 0.95$) (Ihsan *et al.*, 2017). These conflicting results indicate that more research is needed into the relationship between wellness and training load in hockey. In addition to wellness, the relationship between sRPE and oral contraceptive has also been considered, with no significant difference in sRPE ($p = 0.690$) between those using ($n = 7$) and not using oral contraceptives ($n = 16$) (Crewther *et al.*, 2018). Finally, sRPE was used to evaluate the use of cold-water immersion and passive recovery techniques following hockey competition (Krueger *et al.*, 2019). There was no significant difference in sRPE ($p = 0.58$) in subsequent matches between groups, suggesting that cold-water immersion does not reduce future exertion levels compared to passive recovery techniques (Krueger *et al.*, 2019).

4.4.3 Limitations

A large limitation of this review is the risk of bias across studies in terms of publication bias, with data only being available on specific hockey populations. Although not directly measured, the results indicate significant risk of bias across studies due to the varied implementation of athlete monitoring and publication bias. Specifically, there was significantly more research available on elite hockey than sub-elite hockey, despite there being many more hockey players competing at the sub-elite level. This discrepancy is likely a result of additionally monitoring in elite populations due to increased resources and interest in maximizing performance at the highest levels. Furthermore, only studies written in the English language were eligible for inclusion in this review, and, a result, the majority of studies were performed on hockey players in English speaking countries. Finally, although meta-analyses were performed for total distance and workrate, and a large number of player-match files were analyzed, the number of studies included was small, particularly when positional breakdowns were considered. Therefore, care should be taken when generalizing the results of this review and meta-analysis to other hockey populations.

4.4.4 Future Directions

Further research across hockey populations is needed to address the gaps in the literature identified in this review. Most significantly, there is a lack of clear definitions for both internal and external training load variables, limiting comparisons across studies and hockey populations. A standardized set of speed, acceleration, deceleration, and heart rate zone thresholds need to be established and implemented in future research. Although individualized thresholds and varied definitions could be used when warranted by the research questions, providing standardized data in appendices would allow for comparisons between hockey populations and meta-analyses of other training load variables. The author recommends that the speed zones of 0-8.0, 8.1-12.0, 12.1-16.0, 16.1-20.0, and >20.0 $km{\cdot}h^{-1}$ be adopted as the standardized thresholds, with these zones generally representing the movement categories of walking, jogging, running, fast running, and sprinting. Although there is little consensus in the literature, these zones align with the work of McGuinness *et al.* (2019) and Vescovi (Vescovi and Frayne, 2015; Vescovi, 2016; Vescovi and Klas, 2018), and the 16.0 $km{\cdot}h^{-1}$ threshold provides a similar threshold to the 15 or 15.5 $km{\cdot}h^{-1}$ cutoffs often used to denote high speed running (Jennings *et al.*, 2012a; Jennings *et al.*, 2012c; Liu *et al.*, 2013; Ihsan *et al.*, 2017; Sunderland and Edwards, 2017; Casamichana *et al.*, 2018; Polglaze *et al.*, 2018; Morencos *et al.*, 2019). In addition to standardized thresholds, future research should also examine the relationship between internal and external training load metrics. Although some studies presented both internal and external load data, the relationships between these measures were rarely considered. Determining standardized relationships between internal and external training load in various hockey populations would allow practitioners to identify when athletes are struggling, perhaps due to overtraining, fitness levels or psychological strain, with disproportionately high internal training loads for a given external load. Finally, most of the research in hockey has been performed in a tournament or test series scenario, likely due to the majority of international hockey taking place in these settings. However, as international hockey is now also played in a league setup, with the implementation of the International Hockey Federation Pro League in 2019, future research will be needed to consider international hockey in a league format.

4.5 Conclusion

This study quantified internal and external training load in male and female hockey and provided confidence intervals for total distance and workrate in elite hockey, both overall and by position. Significant differences in total distance were found between male and female hockey with male athletes covering about one additional kilometer per match. Although frequently measured, speed zones thresholds differed between studies, preventing direct comparisons. Acceleration, deceleration, and maximal sprint speed were less frequently measured. Compared with external training load, fewer studies have measured internal training load in hockey with the majority of those doing so focusing on heart rate based measures and sRPE. The summary data provided in this study were limited by the hockey populations on which research has been performed. Future research should focus on the development and use of standardized definitions for training load variables and thresholds to allow for further analysis across hockey populations. The data provided here can be used tailor training to ensure that athletes are appropriately preparing for the demands of competition, thereby minimizing the risk of injury and improving performance.

Chapter 5: The Relationship between Athlete Recovery and Training Load in Hockey

As outlined in the previous chapter, the vast majority of published data on athlete monitoring has focused on measuring performance and load in training or competition settings (Saw, Main and Gastin, 2016; Taylor *et al.*, 2017; Fox *et al.*, 2018). However, even amongst elite performers, athletes only spend a relatively small percentage of their total time completing sports-specific activities (Sperlich and Holmberg, 2017). If athlete monitoring solely focuses on measuring match-demand variables such as training load, which only provide information on the on-pitch environment, coaches and sports-scientists will not be able to accurately determine the recovery status and wellbeing of their athletes (Nässi *et al.*, 2017a; Kraft *et al.*, 2018). Therefore, this study will evaluate those monitoring metrics not considered in the systematic review that allow practitioners to understand how athletes are responding to the match and training-demands of hockey. Unfortunately, due to the impact of COVID-19 (please see Appendix C), this study was significantly modified and shortened as a result of national lockdowns and government restrictions, with no post-testing performed. However, the data collected still provide a framework to indicate which recovery measures may be appropriate for evaluating athlete recovery status in hockey, to determine if athletes are responding well to the training and match-demands.

5.1 Rationale

Recovery monitoring refers to athlete monitoring techniques focused on athletes' wellbeing and response to a given load, outside of a training or competition environment (Saw, Main and Gastin, 2015; Taylor *et al.*, 2017; Fox *et al.*, 2018). The research of Sperlich and Holmberg (2017) demonstrates that the choices that athletes make throughout the day, such as decisions regarding nutrition, sleep, and activity levels, as well as other roles, responsibilities and stressors in their lives, impact athletes' physical and psychological response to training. In order to achieve maximal benefits from training, a balance must be maintained between training stress and recovery (Kellmann, 2010). Although the principle of progressive overload indicates that successful training programs require overload, this overload is only beneficial when accompanied with adequate recovery (Lambert and Borresen, 2006; Meeusen *et al.*, 2013; Duffield *et al.*, 2018). When excessive training is continued for extended periods of time, without sufficient recovery, athletes enter a maladaptive state of recovery-stress imbalance, causing extreme fatigue, overreaching, eventually overtraining (Nässi *et al.*, 2017a). Athletes in these maladaptive states will experience a sport-specific drop in performance paired with reduced wellbeing and a disturbance in mood state, for which there is no treatment apart from adequate rest (Meeusen *et al.*, 2013). As Kellmann *et al.*, (2018) identified, the aim of recovery monitoring is to ensure that athletes maintain a balance between stress and recovery to promote overall wellbeing and prevent a maladaptive training state.

As athletes respond differently to identical training programs, recovery monitoring is critical to provide information on how each individual is responding to training (Kölling *et al.*, 2015). The same training load that results in positive adaptation in one athlete may cause overreaching and or overtraining in another so even the most carefully planned program will not be best suited to all members of a team (Lambert and Borresen, 2006; Kölling *et al.*, 2015). If athlete recovery and wellbeing are not monitored, it is often impossible to know when an individual athlete is experiencing training distress until notable performance decrements, often in the form of repeated stress injuries or illness have occurred, at which point recovery may take weeks or months (Meeusen *et al.*, 2013). Therefore, regular recovery monitoring is important to ensure that individual instances of under-recovery are identified and managed before overtraining occurs (Lambert and Borresen, 2006; Kellmann, 2010).

This study will monitor athlete recovery using both the subjective RESTQ-S and the objective CMJ (Kallus and Kellmann, 2016). As subjective and objective approaches both have their strengths and limitations, a multifaceted protocol has been recommended (Saw, Main and Gastin, 2016; Duffield *et al.*, 2018). A validated questionnaire, RESTQ-S incorporates information on the physical, behavioral, social, and subjective aspects of both stress and recovery (Kallus and Kellmann, 2016). It consists of 76 questions on a 7-point Likert scale making up 19 categories - 10 for stress and 9 for recovery, 12 of which are generic and 7 of which are specific to sport (Martinent *et al.*, 2014; Kallus and Kellmann, 2016). RESTQ-S has been shown to be an effective method of measuring athletes' responses to training in a variety of settings across sports (Kellmann and Klaus-Dietrich, 2000; Kellmann *et al.*, 2001; Jurimae *et al.*, 2002; Coutts, Wallace and Slattery, 2007; Freitas *et al.*, 2014; Nunes *et al.*, 2014; Saw, Main and Gastin, 2016;

Nässi *et al.*, 2017a). In addition, as discussed in section 3.4.1, maximal jump height, a measure of neuromuscular fatigue, has been repeatedly shown to be a valid and reliable indicator of athlete recovery and fatigue (Chambers *et al.*, 1998; Coutts *et al.*, 2007; Andersson *et al.*, 2008; Delextrat, Trochym and Calleja-Gonzalez, 2012; Johnston *et al.*, 2013; Wiewelhove *et al.*, 2015; Nässi *et al.*, 2017a; Starling *et al.*, 2019).

Despite the importance of recovery monitoring, few studies had considered recovery monitoring in hockey populations prior to late-2019 when this study was developed (Parrado *et al.*, 2010; Kölling *et al.*, 2015; McGuinness *et al.*, 2018; Ihsan *et al.*, 2017; Vescovi, Klas and Mandic, 2019). Recovery monitoring is of particular importance in hockey because the unlimited rolling substitutions allow hockey to be played at a higher intensity than other field-based team sports, with athletes averaging 85-89% of their maximum heart rate while on the pitch (Lythe, 2008; Sell and Ledesma, 2016; Vescovi, 2016; McGuinness *et al.*, 2017). Additionally, the physical demands of the game vary by playing position, with average differences of up to 2.3 km per match, so athletes in different positions may have notably varied responses to a given training stimulus (Gabbett, 2010; Boran, 2012; Vescovi, 2016; McGuinness *et al.*, 2017; Sunderland and Edwards, 2017). Where recovery monitoring has been studied in hockey, most of the research has been short in duration (≤ 10 days) and occurred in a tournament or training-camp environment (Parrado *et al.*, 2010; Kölling *et al.*, 2015; Ihsan *et al.*, 2017; McGuinness *et al.*, 2018; Vescovi, Klas and Mandic, 2019; McMahon, Sharp and Kennedy, 2021; Krueger *et al.*, 2019; Vescovi, 2019). However, as overreaching and overtraining take time to develop, it is important to consider how athlete recovery relates to training load over longer periods of time (Meeusen *et al.*, 2013). To be an effective marker of athlete recovery, a measure should be sensitive to changes in training load and be able to differentiate athletes responding well to training and those entering a maladaptive training state (Meeusen *et al.*, 2013). This study focused on the first of these two requirements by examining the responsiveness of RESTQ-S and CMJ height to changes in internal and external training load. The aim of this study was to determine the dose-response relationship between training load and athlete recovery status, as measured by RESTQ-S and CMJ height, over the course of an elite hockey season.

5.2 Methods

A natural experiment was performed, with repeated measures taken for a period of four weeks during a competitive hockey season. The outcome variables were athlete recovery status, as measured via RESTQ-S and CMJ height, and external and internal training load, monitored via GNSS units and heart rate monitors. Athlete recovery status was assessed on a weekly basis, and training load monitoring occurred during all tri-weekly training sessions and weekly competitions.

5.2.1 Participants

This study began with 17 female participants from Durham University Hockey Club's first team. As a match-day squad consists of 15 outfield athletes, this sample size was selected to allow for monitoring of all outfield athletes regularly competing at the first team level. However, due to COVID-19 restrictions, many participants had to complete a period of self-isolation during the four-week testing period. To be included in the final analysis, athletes were required to have participated in at least three weeks of the study. As a result, the final sample consisted of ten athletes (22.3 ± 2.6 years, 167 ± 4 cm, 62 ± 6 kg). All participants competed in England Hockey's Division I North (one tier below the premier league) and had an average of 12 ± 4 years of experience playing hockey. Goalkeepers were excluded from this research due to the difference in the demands of their position. Prior to the start of the study, all participants completed a prescreening questionnaire to ensure that they were free from serious injury and were not at an elevated risk of cardiovascular complications from exercise (Appendix B). Ethical approval was obtained from the university ethics board, and all participants were required to provide informed consent (Appendix D). All relevant government, university, and England Hockey guidelines were followed in relation to COVID-19 protocols operating at the time of the study.

5.2.2 Procedures

5.2.2.1 Recovery monitoring

Athlete recovery monitoring consisted of the RESTQ-S questionnaire and a CMJ testing occurring weekly. To control for confounding variables, participants were asked to complete the RESTQ-S questionnaire each Monday morning within one hour of waking. The questionnaire was completed via a secure online google form. To promote compliance, athletes who failed to respond to the questionnaire by mid-morning on Monday received a prompt, reminding them to complete the questionnaire as soon as possible. Athletes were asked not to discuss their responses with others to reduce the effect of peer pressure. RESTQ-S responses were used to calculate scores for the four overall scales of general stress, general recovery, sport stress, and sport recovery (Kallus and Kellmann, 2016). Specifically, scores for each of the seventeen individual subscales were taken as the sum of the responses from the four corresponding questions, and the overall scale scores were calculated as the mean of the relevant individual subscales (Kallus and Kellmann, 2016). The overall scales were chosen in favor of the seventeen individual subscales in order to minimize type I errors, given the limited dataset. This approach is in alignment with the overall two-factor structure of the questionnaire (Kallus and Kellmann, 2016).

Countermovement jump testing took place prior to athletes' training sessions on Monday evenings. This testing was conducted at approximately the same time each week to minimize the impact of circadian variation (Atkinson and Reilly, 1996), and athletes were familiarized with vertical jumping prior to data collection, minimizing any learning effect. Jump height was measured using an Optojump optical measuring system (Microgate, Bolzano, Italy), which has been shown to be a valid measure of vertical jump height (ICC = 0.997-0.998) (Glatthorn *et al.*, 2011). Athletes were instructed to keep their hands on their hips during the jump to remove the effect of arm swing, minimize the learning effect, and isolate lower-limb power (Heishman *et al.*, 2020). Each athlete was given three attempts per week, the highest of which was recorded.

5.2.2.2 Fitness testing

Athletes' aerobic conditioning was assessed at the start of the study. Testing was planned for the end of the study; however, due to a coronavirus lockdown, post-testing could not take place. Fitness testing consisted of the 30-15 intermittent fitness test, a maximal on-field test (Buchheit, 2010). The test, which has been shown to be valid and reliable (ICC = 0.91, correlation with velocity of maximum oxygen consumption during a continuous aerobic assessment $r = 0.67$, $p = 0.013$), requires athletes to run for 30 seconds followed by 15 seconds of active recovery, beginning at 8 $km\cdot hr^{-1}$ and increasing by 0.5 $km\cdot hr^{-1}$ until voluntary exhaustion or completion of 30s at 24.5 $km\cdot hr^{-1}$ (Buchheit *et al.*, 2009; Covic *et al.*, 2016; Bruce and Moule, 2017; Valladares-Rodríguez *et al.*, 2017). The test takes place on a 40m course, with speed and timing dictated by an audio file (Buchheit, 2010). The 30-15 test has good external validity, performed on a hockey pitch and including change of direction, as occurs during match play. To enhance the validity and reliability of the fitness test data, a familiarization procedure was conducted for any participant who had not previously completed the 30-15 test. Participants were asked to abstain from alcohol and strenuous training for 24 hours prior to testing. Participants wore heart rate monitors during the fitness test (Firstbeat Sports, Firstbeat Technologies Oy, Jyvaskyla, Finland), and the maximal heart rate recorded for each individual during the test was taken as their maximum heart rate, unless they exceeded this heart rate during a hockey session (Buchheit, 2010). Athletes' scores were recorded as the speed of the final 30-second interval completed, with an additional 0.25 added to their score (half of a level) if they completed at least half of the next 30s stage before falling behind the specified pace.

5.2.2.3 Training load measurement

Training load measurement occurred during participants' regularly scheduled training and competition for the duration of the study. External training load was measured using Catapult's Vector S7 GNSS devices (Catapult Vector S7, Catapult Sports, Melbourne, Australia), which were tested and shown to be a valid and reliable measure of hockey-specific movement patterns, with a correction factor of 1.0286 used to adjust for the negative bias of 2.78% identified in these units (Chapter 6). The GNSS monitors were worn between the scapulae in the pocket of a specially formatted vest. External training load was measured in terms of high-speed (>15 $km \cdot hr^{-1}$) running distance (HSR). The 15 $km \cdot hr^{-1}$ cutoff has been recommended for the monitoring of elite hockey athletes (Hamilton, 2019) and aligns with cutoffs used by many studies in the literature (Jennings *et al.*, 2012b; Jennings *et al.*, 2012c; Ihsan *et al.*, 2017; Sunderland and Edwards, 2017; Polglaze *et al.*, 2018; Morencos *et al.*, 2019). Other external training load measures initially considered include total distance (TD), distance in other speed zones, equivalent distance and workrate ($m \cdot min^{-1}$). However, due to the limited sample size, only one measure of external training load was chosen for the dose-response analysis to minimize the number of comparisons and decrease the likelihood of type 1 error. HSR was selected instead of TD to provide a measure of high-intensity work and to minimize collinearity, as TD has been shown to more strongly correlate with training impulse, the measure of internal training load used in this analysis (Konerth, 2019).

Internal training load was monitored via heart rate using Polar H1 heart rate monitors (Polar Electro Oy, Kempele, Finland) downloaded via the Vector S7 units. The primary outcome variable was training impulse (TRIMP), calculated using a modified female training impulse algorithm (fTRIMP) (Konerth, 2019). This algorithm was developed by the author in the work directly preceding this thesis. Female TRIMP is based on the relationship of blood lactate and exercising heart rate during submaximal exercise determined from a similar group of female university hockey athletes (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019). Female TRIMP was selected as the summary variable for internal training load as it better summarizes the demands of intermittent sports than average heart rate and avoids the subjective nature of rating of perceived exertion (Konerth, 2019). As with external load data, heart rate data was analyzed from the beginning of warmup until the completion of the cooldown period. Athletes wore the same heart rate and GNSS monitors for the duration of the study.

In order to accurately determine athlete training load status at the time of recovery monitoring evaluation, an exponentially weighted moving average (EWMA) approach was utilized (Sampson, Fullagar and Murray, 2017). The EWMA avoids the limitation of rolling averages, which weights all days across a measurement period equally regardless of their recency (Sampson, Fullagar and Murray, 2017). By placing more weight on the latest training sessions, the EWMA approach accounts for the decaying nature of a training load's impact on fitness and fatigue level. Additionally, when used to calculate acute:chronic workload ratios, the EWMA approach has been shown to be a more sensitive predictor of athlete injury than ratios calculated via rolling averages (Murray *et al.*, 2017). For these calculations, a decay constant of seven was chosen due to the weekly nature of the recovery monitoring protocol. Therefore, daily load ($EWMA_{today}$) was calculated as follows, with $Load_{today}$ being the training load recorded for the athlete on a given day and $EWMA_{yesterday}$ being the exponentially weighted training load value calculated for the day prior (Murray *et al.*, 2017).

$$EWMA_{today} = (Load_{today} \times 0.25) + (EWMA_{yesterday} \times 0.75)$$

In order to provide an accurate calculation of load on the first day of the study, a combination of preliminary data collected in the week prior to data collection (week 0) and data from the first week of the study (week 1) was used to approximate athlete training load in week 1. As week 0 and week 1 were designed by the coaching staff to be similar in terms of athlete load and volume, this provided an accurate approximation of training load, which was then used to determine $EWMA_{yesterday}$ on the first day of data collection. As recovery monitoring evaluation occurred prior to training each Monday (days 8, 15, 22, and 29), athlete training load values were taken as their EWMA load on Sunday (days 7, 14, 21, and 28). Average weekly training load was also reported to allow for comparisons across studies; however, EWMA values were used for all dose-response analyses.

5.2.3 Statistical analysis

		Week 1	Week 2	Week 3	Week 4	Overall
Training Load	Training Impulse (AU)	879 ± 210	806 ± 178	1026 ± 255	142 ± 47	777 ± 356
	High Speed Running (m)	2911 ± 713	3091 ± 967	3441 ± 957	398 ± 153	2690 ± 1316
	Total Distance (m)	24678 ± 2526	21627 ± 3239	27932 ± 5010	5668 ± 2197	21567 ± 8316
Athlete Recovery	Sport Stress (AU)	7.6 ± 3.3	7.3 ± 3.5	9.9 ± 6.1	5.7 ± 4.0	7.7 ± 4.2
	Sport Recovery (AU)	11.8 ± 3.0	12.8 ± 2.9	12.0 ± 3.1	12.0 ± 3.6	12.2 ± 3.0
	General Stress (AU)	8.6 ± 2.7	8.5 ± 2.9	9.8 ± 3.1	8.3 ± 2.7	8.8 ± 2.8
	General Recovery (AU)	12.9 ± 2.5	14.0 ± 2.5	12.4 ± 9.9	13.2 ± 2.3	13.2 ± 2.4
	CMJ Height (cm)	31.7 ± 5.3	31.7 ± 6.0	35.2 ± 3.9	32.5 ± 7.9	32.2 ± 5.8
Sample Size	Training load (n)	10	10	10	6	36
	RESTQ-S (n)	10	10	7	6	29
	CMJ Height (n)	10	10	3	6	33

Table 5.1: Average Weekly Training Load and Athlete Recovery

Training load data were downloaded and analyzed via Catapult's OpenField software (Catapult Sports, Version 2.5.0, Melbourne, Australia). RESTQ-S data were collected via a secure online form, and results were combined to provide scores for each of the subscales and combined scales, as has been previously validated (Martinent *et al.*, 2014; Kallus and Kellmann, 2016). All data were collated in Microsoft Excel (Microsoft Corporation, Version 2002, Redmond, Washington), and statistical analyses were performed in R (R Core Team, 2021). Data were checked for normality using visual inspection of Q-Q plots and histograms. Regression analyses were performed via a repeated measures correlation (Bland and Altman, 1995) using the Rmcorr package (Bakdash and Marusich, 2017; Bakdash and Marusich, 2021). In accordance with the increasing evidence against the use of traditional null hypothesis significance test p-values

(Wasserstein, Schirm and Lazar, 2019), p-values for the correlation coefficients of the repeated measures correlation were calculated via the minimum effects test (Lakens, Scheel and Isager, 2018). The range of (-0.1, 0.1), for which correlation coefficients are considered to be trivial, was taken as the range of no practical significance. Two one-sided tests were performed using Laken's spreadsheet for equivalence testing (Lakens, 2017), with the p-value for the minimum effects tests calculated as the inverse (1-x) of the result of the equivalence testing (Lakens, Scheel and Isager, 2018). Regression data are presented alongside 95% confidence intervals, and statistical significance was set to $p < 0.05$.

5.3 Results

Participants' average score on the 30-15 intermittent fitness test was 19.6 ± 1.2 km·hr^{-1}. Athletes' average weekly training load and recovery scores are summarized in Table 5.1, alongside the sample size for each week. Athlete training load values were similar across the first two weeks, peaked during the third week, and dropped-off in the final week. This decrease in the fourth week was the result of an intentional in-season deload period. Excluding the deload week, athletes covered an average of 24746 ± 4459 m per week, 3148 ± 885 m of which were at speeds greater than 15 km·hr^{-1}, resulting in a weekly internal load (fTRIMP) of 904 ± 229 arbitrary units (AU). Internal and external training load measures were very strongly correlated with a repeated measures correlation of $r = 0.91$ (0.88 – 0.93) between TD and fTRIMP and $r = 0.85$ (0.81 – 0.88) between HSR and fTRIMP. External training load measures were also highly correlated with a correlation of $r = 0.91$ (0.88 – 0.93) between HSR and TD. EWMA training load values for days 7, 14, 21, and 28 are shown in Figure 5.1. As expected, these values differ from but reflect a similar trend to average weekly values.

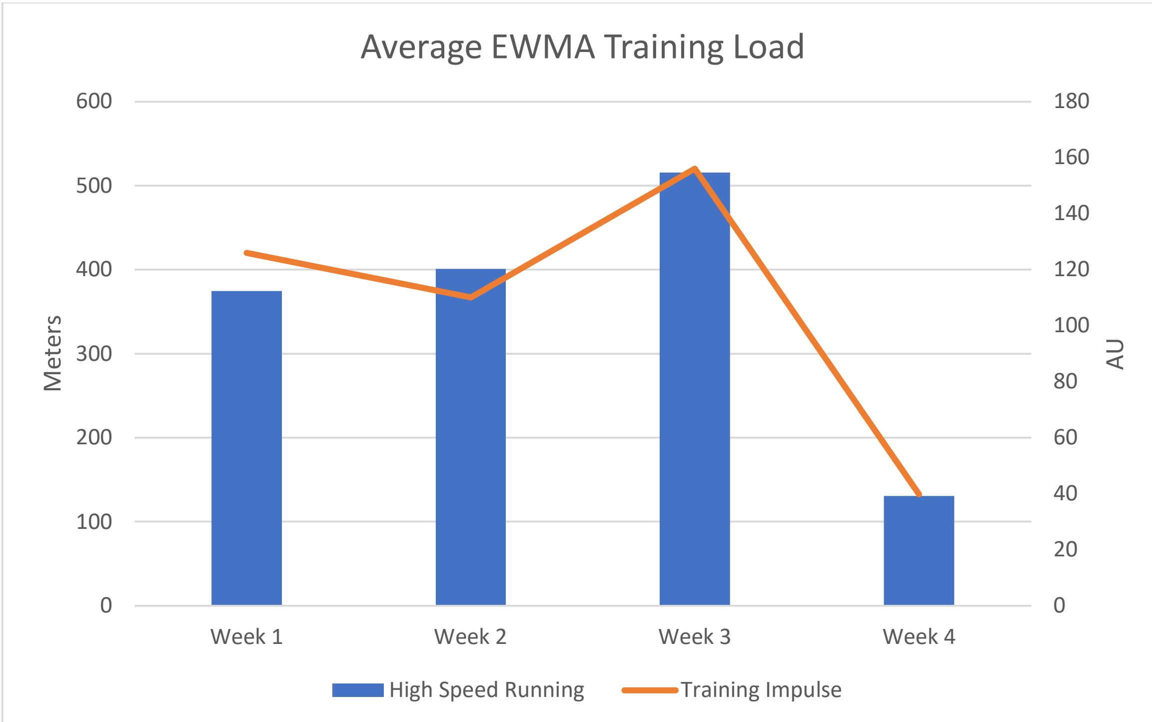


Figure 5.1: Exponentially Weighted Moving Averages (EWMA) Training Load

Athlete recovery measures varied less notably over the four weeks than training load. Sport stress and general stress followed a similar pattern to training load with comparable values across the first two weeks, a peak in the third, and decrease in the fourth week. However, sport recovery peaked in the third week and general recovery peaked in the second, rather than in the final, deloaded week. Excluding the deload week, average weekly values were 8.1 ± 4.2 AU for sport stress, 12.2 ± 2.9 AU for sport recovery, 8.9 ± 2.8 AU for general stress, 13.2 ± 2.5 AU for general recovery, and 32.1 ± 5.4 cm for CMJ height.

Table 5.2: Dose-Response Relationship between Training Load and Athlete Recovery

		Training Impulse			**High Speed Running**		
Athlete Recovery		r	CI	p*	r	CI	p*
RESTQ-S	Sport Stress	0.57	(0.19, 0.80)	0.006	0.51	(0.12, 0.77)	0.010
	Sport Recovery	-0.08	(-0.48, 0.35)	0.542	-0.05	(-0.46, 0.38)	0.591
	General Stress	0.47	(0.06, 0.74)	0.030	0.48	(0.07, 0.75)	0.027
	General Recovery	-0.31	(-0.65, 0.13)	0.156	-0.34	(-0.67, 0.10)	0.123
Countermovement Jump		-0.09	(-0.54, 0.39)	0.506	-0.06	(-0.51, 0.42)	0.568

**Calculated via a minimum effects test*

Four athletes had to undergo a period of self-isolation in accordance with coronavirus restrictions (in operation at the time) during the last week of the study so were excluded from week 4. As a result, the sample size dropped from ten athletes to six for the final week, and these four athletes were also not able to complete vertical jump testing for week 3. Three other athletes also missed jump testing in week 3 due to outside commitments, further reducing the sample size. Therefore, although CMJ height was notably higher in week 3, this result is arbitrary as the increase was likely due to the athletes who were available for testing having a CMJ height above the team average, rather than individual athletes increasing their jump height. Participant compliance with completing the RESTQ-S was lowest in week 3, with three athletes not completing the questionnaire after receiving prompts to do so.

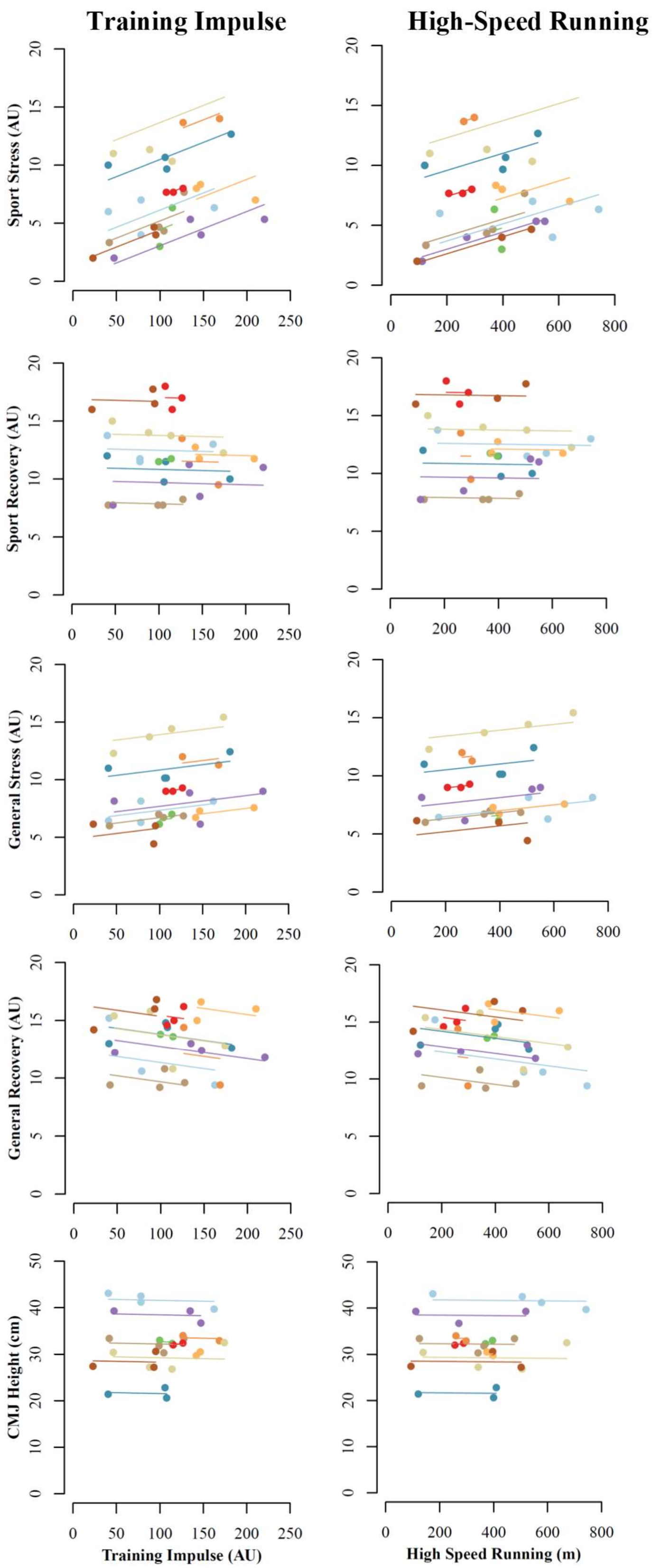


Figure 5.2 Repeated-Measures Correlations between Training Load and Athlete Recovery

The results of the repeated measures correlations between training load and athlete recovery are summarized in Table 5.2 and illustrated in Figure 5.2. The RESTQ-S scale of sport stress was most sensitive to changes in training load, followed by general stress, the correlations between which were all significantly non-trivial by the minimum effects test (p = 0.006 – 0.030). However, due to the limited sample size the confidence intervals were still very wide demonstrating the uncertainly in the exact strength of the positive relationship. In contrast with the stress scales, the recovery scales were not as sensitive to changes in training load, and although there were negative relationships, these were not significantly non-trivial (p = 0.123 – 0.591). Countermovement jump height was not correlated with changes in training load, with individual athlete scores staying relatively consistent across testing (r = -0.06, -0.09, p = 0.506, 0.567).

5.4 Discussion

The key findings of this study were a moderate positive correlation between athlete training load and the RESTQ-S scales of sport stress and general stress, while sport recovery, general recovery and CMJ height were not significantly correlated with training load (Table and Figure 5.2). As sensitivity to change in athlete training load is critical for recovery monitoring measures, these results would support the use of the sport stress and general stress scales of RESTQ-S for recovery monitoring in hockey. However, the lack of a dose-response relationship with athlete training load contraindicates the use of CMJ height for in-season athlete recovery monitoring. Future research will be needed to determine if RESTQ-S scores can be used to distinguish athletes exhibiting positive and maladaptive training responses in order to determine the overall effectiveness of this recovery monitoring measure.

5.4.1 Training load

The weekly training load values for hockey athletes in this study were consistent with those found in other national-level hockey populations (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019). The majority of studies on hockey athletes have measured training load in tournament rather than league settings, with few studies reporting weekly training load in hockey athletes (Chapter 5). Additionally, in some studies where weekly training load values have been reported, different metrics have been used, rendering comparisons meaningless (Perrotta, Held and Warburton, 2017; Perrotta *et al.*, 2019b). However, in eight elite male hockey athletes, competing in a league one level above the athletes in this study, weekly TRIMP was reported to be 826 ± 123 AU, similar to the 904 ± 229 AU reported here (exclusive of the deload week) (Stagno, Thatcher and Van Someren, 2007). Rather than Stagno's TRIMP, TRIMP was calculated in this study using female TRIMP, which was developed based on the same protocol, in the research directly preceding this thesis (Konerth, 2019). Thus, despite the different algorithms, the results are comparable across sexes because the fTRIMP formula corrects for the 30% bias that occurs when Stagno's TRIMP is applied to female athletes (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019). Both weekly internal and external training load were similar to values reported in a comparable group of top university hockey athletes: fTRIMP: 902 ± 110 AU, TD: 23441 ± 1794 m, HSR: 2639 m (Konerth, 2019). Taken together, these results provide good evidence on the weekly in-season training load of hockey athletes competing at the national league level. No other studies have yet reported EWMA training load in hockey, so more research will be needed to establish expected load values using these weightings in hockey athletes.

The associations between internal and external training load measures reported in this study were much higher than those values reported in a recent meta-analysis across sports but comparable to those found in other studies on hockey competition (Polglaze *et al.*, 2015; McLaren *et al.*, 2018; Konerth, 2019; Tuft and Kavaliauskas, 2020). Specifically, a study of elite men's hockey reported a correlation of r = 0.868 between TD and TRIMP (calculated using Polar's propriety algorithm) in matches, similar to the value of r = 0.91 recorded in this study (Polglaze *et al.*, 2015). A study of elite university athletes reported a correlation of r = 0.949 between TD and fTRIMP over the course of a season (Konerth, 2019). In contrast with these findings, a recent study of eleven male hockey athletes during training reported a TD versus TRIMP correlation of r = 0.573; however, load data was phased to only include running-based training, which likely impacted the lower correlation coefficient (Tuft and Kavaliauskas, 2020). Despite the relative similarities within hockey, the internal versus external training load relationships found in this study were notably higher than those reported in a recent meta-analysis across sports (McLaren *et al.*, 2018). Specifically, the meta-analysis reported a pooled correlation between TD and TRIMP of r = 0.74 (0.56 – 0.86), compared to the r = 0.91 (0.88 – 0.93) measured in this study and a pooled correlation between HSR and TRIMP of r = 0.28 (0.10

– 0.45) compared to the r = 0.85 (0.81 – 0.88) found in this study. However, the meta-analysis consisted of studies on rugby league, Australian football, and soccer, suggesting that the internal versus external training load relationship is sport dependent. This sports-dependency is particularly evident in the relationship between HSR and TRIMP, but this result is to be expected given that different sports have varied rest time between high-speed efforts altering the heart-rate response.

5.4.2 The Recovery-Stress Questionnaire for Athletes

The results of this study showed a significantly non-trivial ($r = 0.47 – 0.57$, $p = 0.006 – 0.030$) positive dose-response relationship between the RESTQ-S stress scales and training load, suggesting that RESTQ-S has the potential to be a good measure of athlete recovery in hockey. There was also a negative, but possibly trivial ($r = -0.31, -0.34$, $p = 0.123, 0.156$), dose-response relationship between general recovery and training load, and no apparent relationship between sport recovery and training load ($r = -0.05, -0.08$, $p = 0.542, 0.591$). These results may seem somewhat counterintuitive, as one would likely anticipate training load to have a greater dose-response relationship with the two sport-specific scales, rather than both stress scales. However, these results align with the wealth of literature on overreaching and overtraining, which suggests that training load can impact overall athlete stress, wellbeing, and mood status (Tobar, 2005; Kellmann, 2010; Meeusen *et al.*, 2013; Duffield *et al.*, 2018). Additionally, the weaker correlation with general recovery is likely indicative of the fact that much of an athlete's recovery is impacted by actions taken and decisions made outside of a training environment. As such, one would expect a weaker dose-response relationship between training load and recovery than between training load and stress. Although the dose-response relationship between general recovery and training load was not significant in this study ($p = 0.123 – 0.156$), there was still a small negative correlation between general recovery and both TRIMP ($r = -0.31$) and HSR ($r = -0.34$), suggesting that with a larger sample size this result may be significant. The same cannot be said for sports recovery, with essentially no association between sports recovery and athlete training load ($r = -0.08, -0.05$).

The results of this study align with previous research that has demonstrated RESTQ-S to be an effective measure of athlete training responses across sports (Kellmann and Klaus-Dietrich, 2000; Kellmann *et al.*, 2001; Jurimae *et al.*, 2002; Coutts, Wallace and Slattery, 2007;

Nunes *et al.*, 2014; Nässi *et al.*, 2017a; Freitas *et al.*, 2014). Much of the previous research on RESTQ-S has focused on the seventeen individual subscales, rather than the four composite scales evaluated in this study or an overall score of the difference between cumulative stress and recovery. However, further analysis of the composite scales and overall scores demonstrates there is still a relationship with previous findings. Freitas *et al.*, 2014 demonstated that, after a period of training load intensification in volleyball, *fatigue* was significantly elevated above baseline ($p < 0.05$) and *somatic complaints* was significantly higher than the control group performing normal training ($p < 0.05$) (Freitas *et al.*, 2014). With both *fatigue* and *somatic complaints* contributing to the general stress scale, these results support the conclusion that general stress is the most responsive scale to changes in training load (Freitas *et al.*, 2014). In addition, a study of elite female basketball athletes reported that recovery-stress state (cumulative stress – cumulative recovery) significantly increased ($p < 0.05$) following a phase of deliberate overload and returned to baseline during the subsequent taper (Nunes *et al.*, 2014). These conclusions align with the findings of this study, as the positive relationship of training load with stress and the negative or trivial correlation with recovery would result in a positive correlation between training load and recovery-stress state.

Focusing specifically on hockey, only one study has reported RESTQ-S scores for athlete recovery monitoring (Kölling *et al.*, 2015). RESTQ-S scores were measured at the beginning and end of a 5-day camp for junior international athletes (Kölling *et al.*, 2015). *Physical complaints* and *injury*, which contribute to the general stress and sport stress scales, respectively, both increased significantly ($p < 0.007$, $p < 0.001$), while *physical recovery*, which contributes to general recovery, significantly decreased ($p < 0.047$) (Kölling *et al.*, 2015). Although correlations with training load were not assessed, these changes further reinforce the results of this study with a strong positive relationship between increased training load and both general and sport stress and a notable but weaker negative correlation between training load and general recovery. Furthermore, none of the individual subscales that make up the sports recovery scale significantly changed when training load increased, again demonstrating the lack of a dose-response relationship with this scale (Kölling *et al.*, 2015). Utilizing a general wellness questionnaire rather than RESTQ-S, a study of the Irish men's hockey team during the 2016 Olympics reported a positive correlation between muscle soreness and TD ($r = 0.649$, $p = 0.031$) (McMahon, Sharp and Kennedy, 2021). Although a direct comparison cannot be made with

RESTQ-S, questions on muscle soreness make up the *injury* subscale that contributes to sport stress, suggesting that muscle soreness may be a key contributing factor to the positive relationship between training load and sport stress.

Comparing the dose-response relationship of RESTQ-S with fTRIMP and HSR, there is little difference between values (Table 5.2). The greatest difference in correlations was for sport stress, which had a larger association with fTRIMP ($r = 0.57$) than HSR ($r = 0.51$), perhaps suggesting that sport stress is influenced more by global internal load than high-intensity work output. However, considering the data as a whole, there appears to be no meaningful differences between the correlations of RESTQ-S with fTRIMP and HSR, with the next largest difference being 0.03. Since it is a measure of internal load, one might expect fTRIMP to have a stronger relationship with recovery-stress status than HSR; however, given the collinearity between fTRIMP and HSR ($r = 0.85$), the similarity of the correlations is unsurprising.

Finally, although the dose-response relationships between training load and RESTQ-S are weak to moderate, these values are within the ideal range for recovery measures (Table 5.2). As recovery monitoring is often performed alongside training load monitoring, it is the additional information, not included in training load, that practitioners are looking to capture (Sperlich and Holmberg, 2017). Therefore, if the correlations between training load and the RESTQ-S scales were very strong, the practical usefulness of these measures for recovery monitoring would be limited since they would provide little additional information to the training load data already collected. Training too much, without adequate recovery, is the main contributor to non-functional overreaching and overtraining, so training load measures must still be responsive to changes in training load (Meeusen *et al.*, 2013; Tobar, 2005). As such, when considering the dose-response relationship between training load and athlete recovery, achieving a middle ground is key. The results of this study would suggest that the RESTQ-S scales of sport stress, general stress, and possibly general recovery achieve this balance of providing additional information to training load, while still being sensitive to changes in it. It would appear that sport recovery is not responsive to changes in training load, possibly indicating that that scale is not a valid measure of athlete recovery; however, due to the width of the confidence intervals, more data will be needed to confirm this conclusion. As the dose-response relationship between training load and athlete recovery is insufficient to determine if a recovery monitoring measure can differentiate between athletes experiencing positive and maladaptive training responses, the results of this study cannot conclusively determine the validity of RESTQ-S as a recovery marker in national-level hockey. The correlations between RESTQ-S and training load provide preliminary evidence and reinforce the results of previous research across sports suggesting that RESTQ-S may be a good measure of athlete recovery status.

5.4.3 Countermovement Jump Height

The results of this study show a trivial correlation between training load and CMJ height ($\rho = -0.09, -0.06$, $p = 0.506, 0.568$). This lack of dose-response relationship would suggest that CMJ height is not sensitive to changes in training load and is therefore not a valid marker of athlete recovery in national-level hockey athletes. Although, CMJ height has been frequently used as a measure of neuromuscular fatigue (Chambers *et al.*, 1998; Coutts *et al.*, 2007; Andersson *et al.*, 2008; Delextrat, Trochym and Calleja-Gonzalez, 2012; Johnston *et al.*, 2013; Wiewelhove *et al.*, 2015; Nässi *et al.*, 2017a; Starling *et al.*, 2019), the results of this study are in agreement with other studies that demonstrate a lack of sensitivity of CMJ height to increases in training load (Freitas *et al.*, 2014; Saw, Main and Gastin, 2016; Krueger *et al.*, 2019; Burt *et al.*, 2020). Additionally, a review of training load measures reported that subjective measures such as RESTQ-S are more sensitive to changes in recovery status than objective measures, such as CMJ jump height (Saw, Main and Gastin, 2016). Specific to hockey, CMJ height was not found to be elevated 1-hour, 24-hours, or 48-hours following a match in male hockey athletes (Burt *et al.*, 2020). Additionally, a study of elite male under-18 hockey athletes reported no significant changes in CMJ height over the course of a 5-day training camp ($p > 0.05$), despite perceived stress and perceived recovery (measured via SRSS) significantly increasing and decreasing, respectively ($p < 0.05$) (Krueger *et al.*, 2019). Considering both RESTQ-S and CMJ height, a study in volleyball athletes reported no significant changes in CMJ height ($p < 0.05$) following an 11-day period of intentionally overloaded training, despite significant changes in RESTQ-S scores (Freitas *et al.*, 2014). Overall, these studies reinforce the findings of this research, demonstrating that CMJ height has a decreased sensitivity to increased training load compared to subjective measures. As CMJ height has been shown here, and in previous research (Krueger *et al.*, 2019; Burt *et al.*, 2020), to be insensitive to varied training load in hockey, it should not be used as a marker of athlete recovery.

5.4.4 Strengths and Limitations

There are several key limitations of this study that should be noted. Firstly, the small sample size and the four-week time domain significantly limit the statistical power. Ideally, if additional data were available, a multilevel linear model would have been performed using a random slope and random intercept model; however, due to the limited sample size and data collection period, there was insufficient power to justify this type of analysis. Additionally, had correlations been stronger, the regression coefficients would have been investigated to approximate the impact of a one-unit change in training load on recovery monitoring, but, again, there was insufficient data to provide meaningful estimates. The impact of menstrual cycle was also not directly considered. Notably, 40% of participants reported using hormonal contraceptives, which have been shown to impact cortisol levels but not Profile of Mood Status (POMS) scores in elite female athletes (Crewther *et al.*, 2015; Wikström-Frisén *et al.*, 2016). Athletes were only monitored during on-pitch training sessions, with any outside individual sessions not monitored due to equipment limitations. However, athletes were asked to self-report any outside training performed, with these results demonstrating that few athletes completed additional training of note. From a practical perspective, it is also important to consider that the length of the RESTQ-S might limit the adoption of this questionnaire in non-research settings. Therefore, future research should also consider shorter, more practical questionnaires, such as ARSS and SRSS which have also been shown to be responsive to increased training load in hockey athletes (Kölling *et al.*, 2015; Krueger *et al.*, 2019).

Finally, an important limitation of this study in determining the validity of RESTQ-S and CMJ height as recovery measures was the lack of a performance indicator. To be a useful marker of athlete recovery, a measure should distinguish between athletes positively responding to training and those entering a maladaptive state of nonfunctional overreaching or overtraining (Meeusen *et al.*, 2013). However, as only the dose-response relationships between training load and athlete recovery were considered in this study, it was not possible to determine if the recovery measures evaluated could identify athletes entering a maladaptive state. Therefore, future research into recovery monitoring in hockey should include indicators of athlete performance in order to determine if recovery measures such as RESTQ-S are not only sensitive to changes in training load, but also can be used to distinguish those athletes entering a maladaptive training state. These studies should include a larger sample size, monitored over an increased time domain as athlete maladaptation takes time to develop and increasing the sample size will allow for more certainty in the results.

5.5 Practical Applications

Despite the aforementioned limitations and impact of coronavirus, the results of this study provide valuable insight into training load and recovery monitoring in hockey which can be used to advance athlete monitoring practices. The athletes in this study had mean weekly training load values of TD: 24746 ± 4459 m, HSR: 3148 ± 885 m and fTRIMP: 904 ± 229 AU (exclusive of the deload week), which are comparable to those found in other national-level hockey populations. These values can be used to tailor pre-season training programs to ensure that athletes are adequately prepared for the demands of the hockey season, thereby reducing injuries and maximizing performance. A significantly nontrivial positive dose-response relationship was found between athlete training load and the RESTQ-S scales of sports stress and general stress, while CMJ height showed no notable correlation with training load. These results indicate that RESTQ-S, particularly sport stress and general stress, is a good candidate for recovery monitoring in hockey, while CMJ height is not a valid marker.

Chapter 6: Validity and Interunit Reliability of Catapult Vector 10 Hz Global Navigation Satellite System Units for Assessing Athlete Movement Patterns in Hockey

Whereas the previous chapter has evaluated the response of athletes to the demands of hockey via the recovery aspect of athlete monitoring, this chapter will now consider the monitoring of on-pitch demands. Specifically, external training load monitoring will be evaluated to assess if current measurement methods are valid and reliable. As shown in the systematic review of match-demands in hockey (Chapter 4) and demonstrated in the previous chapter, external training load in hockey is most frequently measured via distance and speed-based metrics collected via GNSS units. As speed and distance are relatively straightforward and intuitively constructed load measures (compared to, for example, training impulse), the question of validity of these measures actually derives from the measurement devices rather than the measures themselves. In a sport such as hockey where athlete movement patterns are stochastic in nature (McGuinness *et al.*, 2017), it is critical to ensure that the external load data provided from GNSS devices in hockey are valid and reliable before using these measures to monitor athletes.

6.1 Introduction

Global Navigation Satellite System (GNSS) receivers, most commonly GPS receivers, are frequently used to track athlete movement patterns and monitor workload in team-sport athletes (Cummins *et al.*, 2013; Scott, Scott and Kelly, 2016; Buchheit and Simpson, 2017; Beato *et al.*, 2016; Beato, Devereux and Stiff, 2018). GNSS technology allows coaches to measure external load by determining the exact physical output of team-sport athletes, which, without advanced monitoring or video-analysis, is otherwise impossible to determine (Scott, Scott and Kelly, 2016). External load data have become a crucial element of athlete management in elite performance sports, with measures such as distance, speed, and accelerations/decelerations used to determine periodization, training intensity, and rest, in order to reduce injury and improve performance (Buchheit and Simpson, 2017; Malone *et al.*, 2017). Specifically, in hockey, monitoring external training load via GPS has become commonplace in many international and elite national-level teams, with external load information used to inform training sessions, substitution patterns, and athlete recovery (Jennings *et al.*, 2012c; Lythe and Kilding, 2013; White and MacFarlane, 2013; Polglaze *et al.*, 2015; Sunderland and Edwards, 2017; McMahon and Kennedy, 2019; Kim, Cha and Park, 2018; Morencos *et al.*, 2018; Vescovi and Klas, 2018).

GNSS receivers are units that determine exact geospatial positioning by connecting with satellites from a global satellite system (Jackson *et al.*, 2018). The first and most well-known GNSS system is GPS, which consists of a network of 24 satellites initially launched by the US government for military applications (Scott, Scott and Kelly, 2016; Jackson *et al.*, 2018). More recently, other satellite systems have been launched such as the Russian GLObal Navigation Satellite System (GLONASS) and the European Unions' Galileo (Gløersen, Kocbach and Gilgien, 2018). Each satellite in these systems contains an atomic clock and transmits information to receivers which, with a minimum of four satellite signals, determine exact location and altitude based on lag time (Scott, Scott and Kelly, 2016). GNSS receivers that work with multiple satellite systems are an improvement on GPS-only receivers, with the access to multiple systems having been shown to increase interunit reliability and likely improve movement tracking precision in team-sport athletes (Jackson *et al.*, 2018).

GNSS units are also categorized by the frequency at which they collect data, with initial units collecting data at 1 Hz (one data point per second), but 5 Hz, 10 Hz, and 15 Hz units have now become commonplace (Scott, Scott and Kelly, 2016; Beato, Devereux and Stiff, 2018). Increased frequency generally results in greater validity and reliability, with the additional information on athlete location improving distance and velocity calculations, particularly when athletes are working at high speeds or with frequent COD (Scott, Scott and Kelly, 2016). In addition, many GNSS units also contain a triaxial accelerometer that measures acceleration in all three planes and, in some devices, this information is used to supplement GNSS data (Scott, Scott and Kelly, 2016). Given that GNSS technology often contributes to coaching decisions regarding athlete fitness and recovery, it is key that the output from these devices is accurate, even when athletes are completing short bouts with frequent COD (Buchheit and Simpson, 2017). Additionally, as comparisons are made across squads of team-sport athletes, the ability of these units to provide consist, unbiased data for all individuals is of critical importance for practitioners.

Catapult Sports is a leading manufacturer of athlete tracking devices, and previous research has demonstrated good levels of validity and reliability of their devices (Portas *et al.*, 2010; Johnston *et al.*, 2012; Johnston *et al.*, 2014; Scott, Scott and Kelly, 2016; Jackson *et al.*, 2018). However, the accuracy of these units has been shown to notably decrease for short distances covered at high speeds (Jennings et al., 2010a). It is important to evaluate new models as they are developed and Catapult's Vector S7 has recently been introduced to the market (Catapult Sports, 2020). The Vector S7 is designed to track athlete movement patterns with accuracy in indoor and outdoor settings and includes a 10 Hz GPS, GLONASS, Satellite Based Augmentation System (SBAS), and Local Positioning System (LPS). To augment the data provided, a 1000Hz 3D accelerometer, 100 Hz 3D gyroscope and 100 Hz magnetometer are also incorporated into the device (Catapult Sports, 2020).

To date, no independent studies have been published on the validity or reliability of Catapult's Vector devices. Therefore, these new devices require field validation and a reliability assessment to determine their ability to measure external training load measures, such as distance and speed. Although GNSS units are regularly used to track athletes and inform coaching decisions in elite and sub-elite hockey (Jennings *et al.*, 2012b; Lythe and Kilding, 2013; White and MacFarlane, 2013; Polglaze *et al.*, 2015; Sunderland and Edwards, 2017; McMahon and Kennedy, 2019; Kim, Cha and Park, 2018; Morencos *et al.*, 2018; Vescovi and Klas, 2018), only one study has previously assessed the validity of GNSS units in a hockey-specific setting (Macleod *et al.*, 2009). Reliability is also important to consider in team sports such as hockey as individual athletes wear different GNSS units, so consistency is needed across measurement devices to allow for comparisons across athletes. As the physical demands of hockey differ from other intermittent ball sports (Cummins *et al.*, 2013), it is important to assess the accuracy of GNSS devices for monitoring athlete movement via distance and speed in hockey. Therefore, the aim of this study was to assess the accuracy and precision of external training load monitoring in hockey using a new GNSS device (Catapult's Vector S7). Specifically, the objectives were to determine the validity and reliability of these units for measuring distance and speed both in an overall match-simulation circuit and during short bouts of high-speed movement with change of direction, as occur in hockey. The results of this study will provide insight on the validity of distance and speed-based measures of external training load for monitoring hockey athletes.

6.2 Methods

6.2.1 Experimental approach to the problem

A repeated measures and validation design were used in this study. In accordance with the protocols of previous research on the validity and reliability of other athlete tracking devices, this study consisted of a measured circuit, designed to replicate the movement patterns of field-sport athletes (Macleod *et al.*, 2009; Jennings *et al.*, 2010a; Jennings *et al.*, 2010b; Johnston *et al.*, 2014; Johnston *et al.*, 2012; Jackson *et al.*, 2018; Beato, Devereux and Stiff, 2018). Specifically, this study used the sport-specific circuit developed by MacLeod et al. to mimic the movement patterns measured in international hockey players during the 2005-2006 English National League hockey season (Macleod *et al.*, 2009). The primary outcomes for validation and reliability were speed and distance, commonly used measures of external training load. These outcomes were measured via the GNSS units and criterion measures (timing gates, a stopwatch, and measuring tape).

6.2.2 Participants

The study included 10 participants (21.3 ± 2.6 years, 173.6 ± 6.5 cm, 69.2 ± 9.6 kg) from a university hockey club's men's and women's first teams (5 male - 21.4± 2.5 years, 178.0 ± 5.5 cm, 76.0 ± 7.4 kg; 5 female - 21.3 ± 3.0 years, 169.2 ± 3.9 cm, 62.4 ± 6.2 kg). Participants competed at the National League level in both the English Hockey League and British University College & Sports League and had played hockey for 13.2 ± 4.0 years. Prior to the study, all participants completed a prescreening questionnaire to ensure that they were free from serious injury and were not at an elevated risk of cardiovascular complications from exercise (Appendix C). Participants were also asked to abstain from alcohol and strenuous activity for 24-hours prior to testing. Ethical approval was granted by the University Research Ethics Sub-Committee, and all participants provided informed consent (Appendix F).

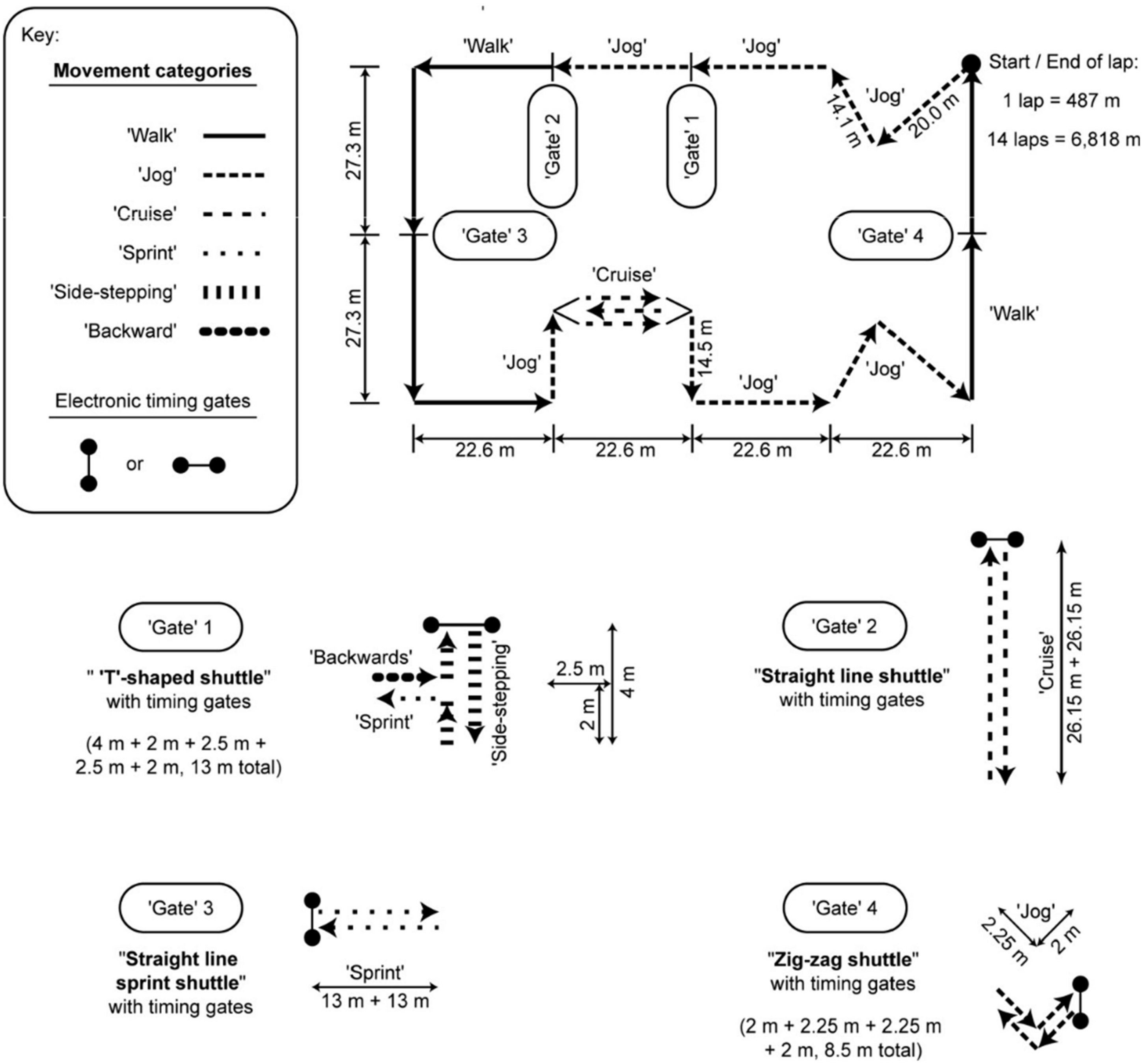


Figure 6.1: Sport Simulation Circuit (Image from MacLeod *et al.,* 2009)

6.2.3 Procedures

As demonstrated in Figure 6.1, the sport-simulation circuit incorporated a mixture of jogging, walking, striding, and sprinting, as well as 4 specific 'gates,' involving lateral and diagonal movement as well as COD (Macleod *et al.*, 2009). Each participant completed 14 laps of the 487 m circuit for a total of 6818m and a duration of approximately 46 minutes (Macleod *et al.*, 2009). All testing took place during an 11-day period on the same hockey pitch and was performed between 8am and 4pm. Distance was determined by a measuring tape, and a stopwatch was used to time each lap. Timing gates (Smartspeed Pro, Fusion Sport, Nottingham, UK) were placed, as indicated in Figure 6.1, to measure the time to complete each gate (Macleod *et al.*, 2009).

In order to determine interunit reliability, each participant wore two Vector S7 units (Catapult Sports, Melbourne, Australia) while completing the circuit. Prior to testing, the GNSS units were turned on and left stationary at the side of the pitch for at least 10 minutes to achieve GNSS lock, as recommended by Catapult Sports (Wilson, 2019). GNSS devices were worn between the scapulae in the pockets of specially designed vests. To prevent any signal interference while the two units were worn simultaneously, the devices were placed adjacent to each other across the back, with the two vests taped and braced to ensure that the units did not overlap (Figure 6.2). The second GNSS unit (units 11-20) was placed in the vest closest to the body and the first unit (units 1-10) was placed on the outside.

Figure 6.2: Placement of GNSS Devices

6.2.4 Statistical analyses

GNSS data were downloaded and analyzed via Catapult's OpenField software (Catapult Sports, Version 2.5.0, Melbourne, Australia). Velocity graphs, positional traces and speed and distance data were used to determine the start point for each lap and gate, to the nearest tenth of a second. Data was collated in Microsoft Excel (Microsoft Corporation, Version 2002, Redmond, Washington) and imported into SPSS for Windows (IBM SPSS, Version 26, Armonk, New York) for statistical analysis.

Data were analyzed by lap, by gate, and for the entire circuit to allow for the consideration of both the overall demands of hockey and short bouts of higher speed action with change of direction. Validity was assessed in terms of bias and standard error of the estimate to determine how accurate speed and distance-based outcome measures are for monitoring external training load in hockey. Additionally, distance and speed reliability were assessed in terms of typical error and interclass correlation to determine the precision of different measurement units. With hockey being a team sport, and individual athletes each wearing a different device, reliability data provides insights as to the precision of measurements across devices, to allow for comparisons across athletes.

Prior to analysis, speed and distance data were checked for normality via visual inspection of Q-Q plots and histograms, with all data found to be approximately normal. Additionally, residual plots of speed data were produced to check for heteroscedasticity, with no cases of heteroscedastic data observed. In order to control for the lack of independence caused by individual athletes wearing two GNSS units, data from only one unit for each individual (randomly selected via coin-toss) was used in validity analyses. Although a repeated-measures approach was taken, with each participant completing 14 laps, each lap was treated as independent for analysis. This approach was selected due to the notable changes in average speed of participants across laps and the identicalness of the course across participants.

Validity for distance measures was assessed both in terms of mean bias, with a one-sample t-test performed, and percent standard error of the estimate %SEE. As no regression analyses could be performed for distance measures, standard error of the estimate, also sometimes referred to as typical error, was defined as the standard deviation of the percent difference between measures and criterion values (Jennings *et al.*, 2010a). Validity for speed was calculated both in terms of mean bias via a paired t-test and linear regression analysis, with the Pearson coefficient (r) and %SEE reported. Reliability data for distance and speed were analyzed via a paired sample t-test, with mean differences reported. Additionally, the typical error (TE), coefficient of variation, and ICC were calculated, as described by Hopkins (Hopkins, 2000), with the quotient of TE and the grand mean used to determine CV (also sometimes referred to as percent TE). In accordance with previous studies and recommendations, typical error was considered to be good (CV < 5%), moderate (CV 5 - 10%), and poor (CV >10%) and Pearson correlations and ICCs were described as trivial (0.0), small (0.1), moderate (0.3), large (0.5), very large (0.7), nearly perfect (0.9) and perfect (1.0) (Macleod *et al.*, 2009; Jennings *et al.*, 2010a; Johnston *et al.*, 2014; Scott, Scott and Kelly, 2016; Beato, Devereux and Stiff, 2018). Significant was set at $p < 0.05$, and data is presented alongside 95% confidence intervals.

6.3 Results

Tables 6.1 and 6.2 outline the results of validity testing for distance and speed, respectively. Percent bias increased as the distance shortened and COD was added, with gate 1 having the highest error followed by 4, 3, and 2. Percent SEE followed a similar pattern as percent bias, with smaller values for the overall course and individual laps, and larger errors for gates 1 and 4. However, %SEE was less than 2.3% for all measures except gates 1 and 4, indicating that although a bias did occur, this bias was relatively consistent, and there was very good agreement in the measurements. This is further evidenced by the near perfect correlations for these speed measures ($r > 0.97$). However, the agreement does notably drop-off for gates 1 and 4 with decreased correlation coefficients and increased %SEE.

Table 6.1: Distance Validity

	GNSS (m)	Criterion (m)	Mean Bias (m)	% Bias	%SEE
Overall	6628.77 (6580.99, 6676.55)	6818	-189.23*** (-237.01, -141.45)	-2.78% (-3.48%, -2.07%)	0.98% (0.67%, 1.79%)
Lap	473.48 (472.55, 474.42)	487	-13.52*** (-14.45, -12.58)	-2.78% (-2.97%, -2.58%)	1.15% (1.03%, 1.30%)
Gate 1	11.14 (11.00, 11.28)	13	-1.86*** (-2.00, -1.72)	-14.32% (-15.38%, -13.23%)	6.46% (5.78%, 7.32%)
Gate 2	50.74 (50.60, 50.87)	52.3	-1.56*** (-1.70, -1.43)	-2.99% (-3.24%, -2.73%)	1.53% (1.37%, 1.73%)
Gate 3	23.88 (23.79, 23.97)	26	-2.12*** (-2.21, -2.02)	-8.14% (-8.50%, -7.79%)	2.13% (1.90%, 2.41%)

Gate 4	7.58 (7.45, 7.70)	8.5	-0.92*** (-1.05, -0.80)	-10.90% (-12.36%, -9.41%)	8.94% (8.00%, 10.13%)

Results presented alongside 95% confidence intervals. %SEE: Percent standard error of the estimate.
Bias is significant at a level of *p < 0.01, **p < 0.05, ***p < 0.001

Table 6.2: Speed Validity

	GNSS ($km \cdot hr^{-1}$)	**Criterion ($km \cdot hr^{-1}$)**	**Mean Bias ($km \cdot hr^{-1}$)**	**% Bias**	**r**	**%SEE**
Overall	8.70 (8.44, 8.97)	8.95 (8.66, 9.25)	-0.25*** (-0.32, -0.18)	-2.79% (-3.57%, 2.01%)	0.975 (0.894, 0.994)	1.09% (0.75%, 2.00%)
Lap	8.73 (8.64, 8.82)	8.98 (8.87, 9.08)	-0.25*** (-0.27, -0.23)	-2.76% (-2.98%, -2.54%)	0.985 (0.979, 0.989)	1.18% (1.06%, 1.34%)
Gate 1	5.99 (5.87, 6.11)	7.14 (6.96, 7.32)	-1.15*** (-1.24, -1.06)	-16.12% (-17.43%, -14.81%)	0.885 (0.842, 0.917)	6.96% (6.22%, 7.91%)
Gate 2	13.83 (13.55, 14.10)	14.27 (13.98, 14.57)	-0.45*** (-0.48, -0.41)	-3.13% (-3.38%, -2.87%)	0.995 (0.993, 0.996)	1.28% 1.14%, 1.46%)
Gate 3	13.36 (13.11, 13.61)	14.60 (14.34, 14.87)	-1.24*** (-1.30, -1.19)	-8.51% (-8.89%, -8.13%)	0.979 (0.971, 0.985)	2.23% (1.99%, 2.52%)
Gate 4	6.32 (6.18, 6.46)	7.19 (6.98, 7.41)	-0.88*** (-1.01, -0.75)	-12.21% (-14.00%, -10.41%)	0.798 (0.727, 0.852)	10.28% (9.18%, 11.69%)

Results presented alongside 95% confidence intervals. r: Pearson coefficient, %SEE: Percent standard error of the estimate. Bias is significant at a level of *p < 0.05, **p < 0.01, ***p < 0.001.

Table 6.3: Distance Reliability

	GNSS First Unit (m)	**GNSS Second Unit (m)**	**Mean Difference (m)**	**% Difference**	**TE (m)**	**CV**	**ICC**
Overall	6638.77 (6593.48, 6684.05)	6619.68 (6568.01, 6671.35)	19.09*** (0.73, 37.45)	0.29% (0.01%, 0.56%)	18.14 (12.48, 33.12)	0.27% (0.19%, 0.50%)	0.926 (0.712, 0.983)
Lap	474.2 (473.30, 475.10)	472.83 (471.84, 473.83)	1.36*** (0.95, 1.78)	0.28% (0.20%, 0.38%)	1.74 (1.56, 1.98)	0.37% (0.33%, 0.42%)	0.907 (0.873, 0.933)
Gate 1	11.10 (10.96, 11.24)	11.07 (10.93, 11.21)	0.03 (-0.01, 0.06)	0.24% (-0.06%, 0.54%)	0.14 (0.12, 0.16)	1.24% (1.11%, 1.41%)	0.971 (0.960, 0.980)
Gate 2	50.79 (50.68, 50.89)	50.74 (50.62, 50.86)	0.04 (-0.06, 0.15)	0.09% (-0.13%, 0.30%)	0.45 (0.40, 0.51)	0.88% (0.79%, 1.00%)	0.559 (0.432, 0.665)
Gate 3	23.91 (23.81, 24.01)	23.78 (23.69, 23.87)	0.13*** (0.06, 0.20)	0.54% (0.26%, 0.82%)	0.28 (0.25, 0.31)	1.16% (1.04%, 1.32%)	0.743 (0.657, 0.810)
Gate 4	7.56 (7.43, 7.68)	7.48 (7.36, 7.60)	0.08** (0.03, 0.13)	1.03% (0.035%, 1.70%)	0.21 (0.19, 0.24)	2.79% (2.49%, 3.17%)	0.915 (0.883, 0.939)

Results presented alongside 95% confidence intervals. TE: Typical error. CV: Coefficient of variation. ICC: interclass correlations. Difference is significant at a level of *p < 0.05, **p < 0.01, ***p < 0.001.

Table 6.4: Speed Reliability

	GNSS First Unit ($km \cdot hr^{-1}$)	GNSS Second Unit ($km \cdot hr^{-1}$)	Mean Difference ($km \cdot hr^{-1}$)	% Difference	TE ($km \cdot hr^{-1}$)	CV	ICC
Overall	8.72 (8.45, 8.98)	8.69 (8.43, 8.94)	0.03 (0.00, 0.05)	0.29% (-0.01%, 0.58%)	0.03 (0.02, 0.05)	0.29% (0.20%, 0.53%)	0.995 (0.977, 0.999)
Lap	8.74 (8.65, 8.83)	8.72 (8.62, 8.81)	0.03*** (0.02, 0.03)	0.29% (0.20%, 0.38%)	0.03 (0.03, 0.04)	0.38% (0.34%, 0.42%)	0.996 (0.995, 0.997)
Gate 1	5.99 (5.87, 6.10)	5.98 (5.86, 6.10)	0.01 (-0.01, 0.03)	0.18% (-0.14%, 0.50%)	0.08 (0.07, 0.09)	1.33% (1.18%, 1.51%)	0.987 0.981, 0.991)
Gate 2	13.84 (13.56, 14.11)	13.83 (13.55, 14.11)	0.01 (-0.02, 0.04)	0.04% (-0.16%, 0.25%)	0.12 (0.11, 0.14)	0.86% (0.77%, 0.98%)	0.995 (0.992, 0.996)
Gate 3	13.39 (13.14, 13.65)	13.32 (13.06, 13.58)	0.07 *** (0.03, 0.11)	0.52% (0.25%, 0.79%)	0.15 (0.13, 0.17)	1.12% (1.00%, 1.27%)	0.990 (0.986, 0.993)
Gate 4	6.33 (6.20, 6.47)	6.27 (6.13, 6.41)	0.07 ** (0.02, 0.11)	1.06% (0.33%, 1.79%)	0.19 (0.17, 0.22)	3.01% (2.69%, 3.42%)	0.946 (0.924, 0.961)

Results presented alongside 95% confidence intervals. TE: Typical error. CV: Coefficient of variation. ICC: interclass correlations. Difference is significant at a level of $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Tables 6.3 and 6.4 illustrate the good level of interunit reliability found in these units. Using qualitative descriptors, the CV is considered ‘good’ in all cases, as values are less than 5%. This high level of validity is further verified by the ICC with very nearly perfect correlations in all cases except for distance measured in gates 2 and 3, where correlations were large and very large, respectively. The GNSS units were connected to an average of 16 satellites during testing (range: 13 - 19).

6.4 Discussion

The aim of this study was to assess the validity and reliability of Catapult Vector S7 units for monitoring the movement patterns of British university hockey athletes competing at the National League level. The sport-simulation circuit mirrored the demands of a hockey match, with time, high speed running distance (15 $km \cdot hr^{-1}$ - 19 $km \cdot hr^{-1}$) and sprint distance (> 19 $km \cdot hr^{-1}$) aligning with values typically recorded by several of these athletes during match play (authors' unpublished data).

The most notable finding of this study was that the Vector units were significantly biased ($p < 0.001$), and consistently underestimated distance by 2.8% overall and up to 14.3% during the 13 m T-shaped gate. As speed is distance over time, there was also a corresponding underestimation of speed with a significant ($p < 0.001$) bias of 2.8% overall and 16.12% during the T-shaped gate. From a practical perspective, the overall bias of 2.8% is relatively small, amounting to a distance of just above two pitch lengths (189 m) over the course of a hockey match, using the distance values suggested by MacLeod et al. (Macleod *et al.*, 2009). Additionally, the precision of this underestimation for both speed and distance, demonstrated by the small %SEE of 0.98% for overall distance suggests that the bias is very consistent, and an adjustment could be made to correct for it.

Table 6.5: Comparison of Distance Measures across Catapult GNSS Models

Device	**Mean Bias (SSC)**	**Mean Bias (short shuttles)**	**%SEE (SSC)**	**%SEE (short shuttles)**
Vector S7 (10 Hz)	2.8% ($p < 0.001$)	8.5m zig-zag jog:14.3% 13m T shuffle: 10.9% 26m shuttle sprint: 8.1% 52.3m shuttle run: 3.0%	1.0%	8.5m zig-zag jog: 8.9% 13m T shuffle: 6.5% 26m shuttle sprint: 1.5% 52.3m shuttle run: 2.1%
MinimaxX S4/v4.0 (10 Hz)	No significant difference from criterion[1]	25.1m sprint with COD: 11.7%[2] 30m straight sprint: 6.5%[3]	2.8%[2]	25.1m sprint with COD: 4.0%[2]
MinimaxX Team 2.5 (5Hz)	No significant difference from criterion[4]	10m zig-zag jog: 7.1%[5] 10m zig-zag stride: 15%[5]	1.5-2.2%[29]	40m zig-zag jog: 9.7%-10.6%[5]

SCC: Sport simulation circuit, %SEE: percent standard error of estimate
[1] (Johnston *et al.*, 2014), [2] (Hoppe *et al.,* 2018), [3] (Castellano *et al.,* 2011), [4] (Johnston *et al.*, 2012), [5] (Jennings *et al.*, 2012), [6] (Morencos *et al.*, 2018).

The measurement bias of 2.8% recorded in this study is notably greater than that reported for many other GNSS units. For example, in McLeod *et al.*'s study of SPI Elite 1 Hz GPS units, from which the sport-simulation circuit used in this study was derived, no significant difference was reported between GPS-measured and actual distance per lap with an error of 0.2 m per lap (0.04%), compared to the error of 13.52 m per lap in this study (Macleod *et al.*, 2009). This finding is contrary to previous results, as 10 Hz models have been repeatedly shown to have greater validity than 1 Hz models (Scott, Scott and Kelly, 2016). Additionally, Catapult performed its own testing on the Vector units and reported a mean bias of 0.8% for total distance over a sport-simulation circuit, which is notably lower than the 2.8% measured here; however, athlete speed was not provided and a lower athlete speed could have contributed to a decreased bias (Catapult Sports, 2019). Table 6.5 demonstrates the how the validity of the Vector S7 units compares to two previous Catapult GPS models, selected for comparison due to the breadth of literature available. Notably, neither of Catapult's MinimaxX units, both precursors to the Vector

S7 model, showed any significant difference from criterion values for distance, with the 10 Hz model having a non-significant mean bias of 0.8% (Johnston *et al.*, 2012; Johnston *et al.*, 2014). Thus, the results of this study indicate a 3.5-fold increase in measurement bias between this device and an earlier unit. However, the error of the Vector models in this study is similar to the 3.17% underestimation bias ($p < 0.001$) measured in the GPSport SPI Pro-X unit (10 Hz receiver with 15 Hz interpolation) on a 13200m multidirectional shuttle course (Rawstorn *et al.*, 2014). The SPI Pro-X devices have been regularly used in recent studies across a range of sports, demonstrating the acceptance of this level of measurement bias (Abian *et al.*, 2015; Cullen *et al.*, 2017; Pueo *et al.*, 2017; Moreno-Pérez *et al.*, 2019; Reddy *et al.*, 2017; Kelly *et al.*, 2020; López-Fernández *et al.*, 2019; Yamamoto *et al.*, 2020).

Although the overall bias of the Vector S7 units was greater than other devices, the mean bias during short, high-intensity movements was similar to that measured in previous Catapult models. As COD increased and overall distance decreased, bias significantly elevated with the largest error observed during the 13m T-shaped gate 1 (14.3%) followed by the 8.5m zig-zag gate 2 (10.9%). Despite their magnitudes, the errors measured on these shuttles are comparable to errors recorded in both Catapult MinimaxX units (Jennings *et al.*, 2010a; Castellano *et al.*, 2011; Hoppe *et al.*, 2017). Notably, there was a much greater error recorded during a zig-zag jogging shuttle by the Vector S7 units than the MinimaxX 2.5; however, this result may have been influenced by added COD, with participants starting and ending in the same location in this study (Jennings *et al.*, 2010a). The mean bias of the Vector unit on the 26 m shuttle sprint (8.1%) is an improvement to that measured by the MinimaxX S4 unit (11.7%) on a 25.1m sprint with COD; however, a lower bias was still recorded on straight 30 m sprint (6.5%), further demonstrating that COD increases bias in GNSS receivers (Hoppe *et al.*, 2017; Castellano *et al.*, 2011).

The results of this study also provide evidence of GNSS error increasing as distance decreases. Specifically, as criterion speed was similar (14.27 v 14.60 $km \cdot hr^{-1}$) between gates 2 and 3, and the only difference in the set-up was the length of the shuttle, the large differences in mean bias and %SEE clearly illustrate the impact of shuttle distance on error. Error significantly increases as distance decreases, as has been noted in previous studies; however, since team sports involve repeated bouts of short high-intensity efforts, these errors likely cause distance and speed to be underestimated in matches (Jennings *et al.*, 2010a; Hoppe *et al.*, 2017). Although each bout may be short in duration, the error will accumulate over the course of a match. However, as this error is common across GNSS models, the bias will be relatively consistent, allowing for comparison across datasets. Additionally, it is worthwhile to note that despite the COD, when the length of the high-intensity effort increased to over 50 m in gate 2, bias decreased to 3% and was not significantly different ($p = 0.21$) than the overall bias recorded. Therefore, although the Vector S7 units, in accordance with other GNSS models, demonstrate an increased bias with increased COD over short distances, this error returns to baseline levels for efforts over 50 m with one COD.

In addition to considering bias, it is also important to note the spread of measurement values, as demonstrated by %SEE. With the exception of gates 1 and 4, speed and distance %SEE ranged from (1.0 - 2.2%), demonstrating good agreement of measurement values, even during the short shuttles in gates 2 and 3. These %SEE are an improvement on the values reported for previous Catapult models on circuits as a whole and similar to values measured for short shuttles with COD (Table 6.5) (Jennings *et al.*, 2010a; Portas *et al.*, 2010; Johnston *et al.*, 2012). Although %SEE notably increased for gates 1 and 4 (6.46%, 8.94%), this result is again likely due to the shortened distance of the shuttles and added change of direction, and is lower than values reported for the MinimaxX 2.5 units (9.7 - 10.6%) (Jennings *et al.*, 2010a). Overall, the low %SEE values indicate that although the speed and distance measures reported by the Vector S7 units were biased, this bias is consistent with a relatively small range of measurement values.

The low CV values (0.27% - 3.01%) and high ICCs (0.559, 0.743, 0.907-0.996) demonstrate the high level of reliability of Catapult Vector S7 units. When the entire circuit was considered, representing the approximate distance covered in a hockey match, there was a mean difference of 0.29% or 19 m and 0.03 $km \cdot hr^{-1}$, values of little practical significance. Notably, this CV of 0.29% is much smaller than the value of 1.4% reported by Catapult during their testing (Catapult Sports, 2019). There were some distances for which mean distance and speed were significantly different between the GPS units; however, this significance was more impacted by the large sample size and precision of the measures than the magnitude of the difference, as the actual percent difference was very small (0.28% - 0.54%) in cases where $p < 0.001$. Although the ICC for distance did decrease for gates 2 and 3, and CV was elevated in gate 4, likely due to increased COD over small distances, the CV was well below the threshold for ‘good’ (5%)

generally accepted in this type of reliability study (Macleod *et al.*, 2009; Jennings *et al.*, 2010a; Johnston *et al.*, 2014; Scott, Scott and Kelly, 2016; Beato, Devereux and Stiff, 2018). Therefore, the results would suggest that Catapult Vector S7 devices are reliable measures of speed and distance.

The reliability of the Vector units is an improvement on the reliability of previous GNSS models. Specifically, the CV for the circuit was 0.27% compared to 1.4 - 3.8% for MinimaxX 2.5 5 Hz, 1.3% for MinimaxX S4 10 Hz, and 1.9% for SPI Pro-X 15 Hz units (Portas *et al.*, 2010; Johnston *et al.*, 2012). Despite the decreased ICC for distance in gates 2 and 3 (0.559, 0.743) and increased CV (2.79%) for distance in gate 4, these values still represent a vast improvement on the reliability of MinimaxX 2.5 5 Hz units over a 40 m zig zag course (CV 7.9 - 10.0%) and a 50 m rectangular course (CV 5.9%) (Jennings *et al.*, 2010a; Portas *et al.*, 2010). Thus, the Vector S7 devices provide reliable speed and distance data for the various movement patterns of hockey.

6.5 Practical Applications

The results of this study indicate that the Vector S7 units are a reliable measure of hockey movement patterns but consistently underestimate overall distance and speed by 2.8%. If these GNSS devices were to be used exclusively, with no comparison to measures from other devices or criterion values, this bias would have little impact as it would remain consistent across sessions and athletes. However, to allow for comparison, an adjustment could be made to correct for the bias, by multiplying results by a factor of 1.0286 (1/ (1 - 0.278)). Since the bias is highly consistent, this correction would provide a very good estimate of actual distance and speed. Although this adjustment would not fully correct for the additional bias measured during the 4 gates, since the direction of the bias is the same, the correction factor would decrease the underestimation. There would still be some error over short distances with COD, so it is important for practitioners to be aware of this limitation, common across GNSS models. The good interunit reliability of these devices indicates that results will be consistent across units. However, it is recommended that the same device is assigned to each athlete across sessions to minimize any small reliability error that could occur. Although the sport-simulation circuit used in this study was specific to hockey, these conclusions can be generalized to other intermittent ball sports with similar physical demands, such as soccer and lacrosse. Overall, the results of this study indicate that Catapult S7 GNSS units are a valid and reliable measure of speed and distance in hockey when the underestimation bias is corrected.

Chapter 7: Developing a pitch-based protocol for calculating individualized training impulse in intermittent field-sport athletes

Complete athlete monitoring requires the consideration of both internal and external training load (Impellizzeri, Marcora and Coutts, 2022). As demonstrated in the previous chapter, external training load metrics often measure clearly defined outputs, and criterion values can be used in assessments of validity (Impellizzeri, Marcora and Coutts, 2022). Conversely, internal training load measures aim to summarize the body's physiological response to exercise through metrics such as heart rate (Impellizzeri, Marcora and Coutts, 2022). Although protocols for the measurement of heart rate data are well established, there is no gold standard method of summarizing these data into internal load scores (Weaving *et al.*, 2017; Fox *et al.*, 2018; McLaren *et al.*, 2018; Passfield *et al.*, 2022). Therefore, this chapter will begin to address the aim of developing an improved procedure for internal training load calculation in hockey athletes. Specifically, the evolution of TRIMP will be considered to determine the most appropriate method for calculating TRIMP in hockey, and a new testing protocol will be developed to increase sport-specificity and external validity.

7.1 An overview of training impulse

7.1.1 Background

Training impulse (TRIMP) is a measure of internal training load derived from an athlete's heart rate (Impellizzeri, Marcora and Coutts, 2022). Heart rate is a suitable marker from which to derive internal training load because it is a well-established, objective method of monitoring exercise intensity and is relatively easy to measure (Banister, 1991). Expressed in arbitrary units, TRIMP incorporates both session duration and heart rate, weighted according to physiological response, to produce a load score summarizing the demands of an exercise session (Banister, 1991). Although the concept and inputs remain the same, various algorithms for calculating TRIMP have been developed over time, with notable algorithms including those developed by Banister (1991), Edwards (1993), Lucia *et al*. (2003), Stagno *et al.* (2007), Manzi *et al.* (2009) and Gonzalez-Fimbres *et al.* (2019).

7.1.2 The evolution of TRIMP algorithms

The concept of TRIMP was first developed by Banister in 1991 to monitor training load in distance runners (Banister, 1991). Incorporating an individuals' average heart rate reserve and session duration, Banister's TRIMP was designed to summarize the physiological demands of steady-state exercise (Banister, 1991). The TRIMP formula weighted heart rate reserve using an exponential curve derived from the relationship between exercising heart rate and blood lactate, the blood lactate vs heart rate reserve (BLvHR) response curve. Banister's TRIMP is calculated as follows:

Equation 7.1: Banister's TRIMP Equation (Banister, 1991)

$$\mathrm{TRIMP} = \mathrm{training\ duration\ (minutes)} \times \mathrm{HRR} \times \mathrm{y}$$

where

Equation 7.2: Banister's TRIMP Sex Weightings (Banister, 1991)

$$\mathrm{y} = 0.64\mathrm{e}^{1.92 \times \mathrm{HRR}} \quad \text{(male)}$$
$$\mathrm{y} = 0.86\mathrm{e}^{1.67 \times \mathrm{HRR}} \quad \text{(female)}$$

and

Equation 7.3: Heart Rate Reserve

$$Heart\ Rate\ Reserve\ (HRR) = \frac{exercising\ heart\ rate - resting\ heart\ rate}{maximum\ heart\ rate - resting\ heart\ rate}.$$

Incorporating exponential models of blood lactate response, Banister's TRIMP algorithm has a physiological basis and avoids giving disproportionate weight to low-intensity activities performed for a long duration. The models are also sex-specific, which is critical, as the unpublished work preceding this thesis illustrated a 30% discrepancy when TRIMP algorithms based upon male athletes were used in female athletes (Konerth, 2019). However, as Banister's TRIMP is calculated from average heart rate over an entire session, it is limited in its ability to monitor intermittent exercise (Stagno, Thatcher and Van Someren, 2007; Gonzalez-Fimbres *et al.*, 2019). Specifically, when heart rate is averaged over intermittent sessions, short, high-intensity efforts have minimal impact on mean heart rate, despite having a notable physiological effect. Therefore, although an effective measure for summarizing continuous exercise where athletes reach steady state, Banister's TRIMP is not an appropriate training load measure for intermittent activities such as team sports.

To overcome this limitation, Edwards developed a new TRIMP algorithm based on summated heart rate zones rather than average heart rate (Edwards, 1993). Edwards' TRIMP is based on five arbitrary heart rate zones (50 - 60%, 60 - 70%, 70 - 80%, 80 - 90%, 90 - 100% max HR) and is calculated by multiplying the time spent in each zone by a weighting factor (1, 2, 3, 4, and 5, respectively) and summing the results (Edwards, 1993). By weighting time spent in each of the heart rate zones, Edwards' TRIMP algorithm prevents short bursts of high intensity from being lost amongst longer periods of low intensity and is therefore better suited for summarizing intermittent activities (Edwards, 1993). However, Edwards' TRIMP represents a simultaneous step forward and step backwards in terms of TRIMP algorithms, for, despite this improvement, Edwards' TRIMP is limited by its generic zones and weightings. Whereas the weightings in Banister's TRIMP algorithm are based on the BLvHR response, Edwards' TRIMP uses general thresholds and weightings without any physiological basis. Although the times in various heart rate zones are incorporated into Edwards' formula, the physiological impact of the times in these zones is likely misrepresented due to the generic zones and weightings.

An improvement on Edwards' TRIMP came a decade later in the development of Lucia's TRIMP (2003), which separates heart rate into zones based on ventilatory thresholds (Lucia *et al.*, 2003). By grouping heart rate according to individuals' ventilatory thresholds, Lucia's TRIMP overcomes the limitation of generic zones in Edwards' TRIMP, with the physiological demands in each of the three heart rate zones theoretically more homogenous than in Edwards model (Lucia *et al.*, 2003). Specifically, a ramp protocol was used to determine the heart rates associated with the ventilatory threshold and respiratory compensation point for each athlete measured via gas-exchange data, with the three zones taken as below ventilatory threshold, between ventilatory threshold and the respiratory compensation point, and above the respiratory compensation point (Lucia *et al.*, 2003). This protocol represents the first individualization of a TRIMP measure, with the results of a fitness assessment used to customize the algorithm for individual athletes. However, despite this advancement, Lucia's TRIMP was still notably limited by the weights used for each of the individualized zones, with integer weights of 1, 2, and

3 used for the three zones, respectively (Lucia *et al.*, 2003). Therefore, despite the physiological basis of the zone thresholds, the generic weightings mean that the physiological load from time in each of the heart zones is not accurately expressed as there is no physiological basis for the demands of the second and third zones being twice and three times that of the first zone.

Progressing beyond many of the limitations of earlier TRIMP algorithms, a modified 'team' TRIMP algorithm with physiologically based heart rate zones and weightings was developed by Stagno *et al.* (2007). To determine the weightings, Stagno *et al.*, returned to the exponential curve that was fundamental to Banister's TRIMP but was notably missing from Edwards' and Lucia's algorithms: BLvHR response (Banister, 1991; Stagno, Thatcher and Van Someren, 2007). Stagno's TRIMP maintains the summated heart rate zones approach of Edwards' and Lucia's TRIMP, thereby avoiding the use of average heart rate, as in Banister's TRIMP. However, rather than generic weightings, the weights for each heart rate zone were taken as the extrapolated blood lactate level for the given heart rate determined by the BLvHR response curve (Stagno, Thatcher and Van Someren, 2007). Rather than using arbitrary zone thresholds, Stagno's heart rate zones 2 and 4 were anchored around heart rate at blood lactate concentrations of 1.5 and 4.0 $mmol \cdot L^{-1}$, representing the lactate threshold and onset of blood lactate accumulation (Stagno, Thatcher and Van Someren, 2007).

Table 7.1: Stagno's TRIMP Zones and Weights (Stagno, Thatcher and Van Someren, 2007)

Zone	% Max HR	Weighting
1	65-71	1.25
2	72-78	1.71
3	79-85	2.54
4	86-92	3.61
5	93-100	5.16

Rather than using equations from the previous published literature, the exponential curve representing the BLvHR response in Stagno's TRIMP was based on empirical evidence. Nine male hockey athletes completed a submaximal lactate threshold test, consisting of four 4-minute stages, starting at 10 $km \cdot hr^{-1}$ and increasing by 2 $km \cdot hr^{-1}$ per stage, with a 1-minute rest between stages (Stagno, Thatcher and Van Someren, 2007). Blood lactate was measured immediately upon completion of each stage, using a fingertip capillary blood sample, with heart rate monitored throughout (Stagno, Thatcher and Van Someren, 2007). Heart rate reserve versus blood lactate was then plotted together for all individuals, with least squares regression performed to calculate the exponential model (Stagno, Thatcher and Van Someren, 2007). Although in many ways Stagno's TRIMP represents a critical step forward in TRIMP modeling for team sport athletes, with physiologically based heart rate zones and weights, its generalizability is limited by both the sample size and the use of only male athletes. The BLvHR relationship differs based on sex, a fact accounted for in Banister's TRIMP with different exponential models used for male and female athletes (Equation 7.2) (Banister, 1991; Konerth, 2019). However, Stagno's TRIMP was developed using a sample of only male athletes (Stagno, Thatcher and Van Someren, 2007). To counter this, an alternative team TRIMP algorithm, termed female TRIMP (fTRIMP), was later developed for female athletes in the unpublished work performed by the author prior to this thesis (Konerth, 2019). Specifically, Stagno's protocol was replicated (with adjustments to the speeds in the lactate threshold test), in 16 female hockey athletes. The resulting values (Table 7.2) are notably different from those for Stagno's TRIMP (Table 7.1), providing further evidence on the need for sex-specific algorithms. However, an analysis of data from across a hockey season was used to determine a very strong linear correlation ($r = 0.996$) between TRIMP calculated using Stagno's TRIMP and female TRIMP in female athletes, with a multiplicative factor of 1.3 able to accurately adjust for sex differences (Konerth, 2019).

Table 7.2: Female TRIMP Zones and Weights (Konerth, 2019)

Zone	**% Max HR**	**Weighting**
1	59-66.9	0.91
2	67-74.9	1.49
3	75-82.9	2.44
4	83-90.9	3.99
5	91-100	6.74

Building on the lactate threshold testing used in the creation of Stagno's TRIMP, Manzi *et al.* developed a fully individualized TRIMP (Manzi *et al.*, 2009). Manzi's TRIMP, termed individualized TRIMP (iTRIMP) is based on an athlete's own BLvHR response curve, rather than group or literary values (Manzi *et al.*, 2009). Even within sexes, metabolic stress can vary amongst athletes exercising at the same heart rate (Manzi *et al.*, 2009; Manzi *et al.*, 2013; Malone and Collins, 2016). Therefore, despite their physiological basis, team TRIMP algorithms such as Stagno's TRIMP and fTRIMP, will vary in the accuracy with which they summarize the physiological demands of exercise across individuals. Integrating individualized BLvHR response curves, iTRIMP overcomes this limitation, producing load scores more reflective of the internal load for each individual athlete (Manzi *et al.*, 2009). To calculate Manzi's iTRIMP for an athlete, he or she must first complete a submaximal lactate threshold test with heart rate monitoring, as described above for Stagno's TRIMP (Stagno, Thatcher and Van Someren, 2007; Manzi *et al.*, 2009). However, rather than combining data across individuals, an athlete's own data is used in the least-squares regression to produce a curve of the form $y = ae^{bx}$ where *a* and *b* are constants, unique to each individual (Manzi *et al.*, 2009). Individualized TRIMP for a session is then calculated as

Equation 7.4: Individualized TRIMP

$$iTRIMP = \frac{1}{n}\sum_{HRR} HRR \times ae^{b\,\times HRR}$$

where *n* is the number of heart rate readings per minute, and HRR is heart rate reserve (Equation 7.3) (Manzi *et al.*, 2009). In addition to individualization, another advancement in iTRIMP is the use of a continuous rather than a stepwise approach to weightings. In iTRIMP calculations, this is achieved by summing over individually weighted HRR values rather than separating heart rate values into discrete zones (Manzi *et al.*, 2009). This increases the accuracy of iTRIMP, as the stepwise approach assumes the physiological demands at the lower and upper edges of the zones are the same, thereby skewing results close to the threshold values (Gonzalez-Fimbres *et al.*, 2019). Using an individualized, continuous approach, iTRIMP is considered to be the best and most physiological accurate TRIMP measure (Manzi *et al.*, 2009; Manzi *et al.*, 2013; Sanders *et al.*, 2017; Fox *et al.*, 2018; Malone *et al.*, 2018).

Although not a direct progression on iTRIMP, it is worthwhile to note the work of Gonzalez-Fimbres *et al.* and their modified TRIMP algorithm (Gonzalez-Fimbres *et al.*, 2019). Gonzalez-Fimbres' TRIMP combines features of both the original and most sophisticated TRIMP algorithms, Banister's TRIMP and Manzi's iTRIMP, into a continuous formula based on established values. Weighting factors are determined by the exponential curves in Banister's TRIMP algorithm (Equation 7.2), rather than an individualized approach (Banister, 1991; Gonzalez-Fimbres *et al.*, 2019). However, instead of using average heart rate, HRR values are individually weighted and summed over time, as in Manzi's iTRIMP (Manzi *et al.*, 2009; Gonzalez-Fimbres *et al.*, 2019). Thus, Gonzalez-Fimbres' TRIMP preserves the benefits of a continuous, physiologically based approach to weightings, without the individualization. Although this algorithm does not account for the variation in metabolic response of athletes exercising at the same HRR, it maintains sex-specificity and can be used to approximate internal load when an individuals' BLvHR response curve is unknown. Therefore, Gonzalez-Fimbres' TRIMP provides a useful progression on Stagno's TRIMP and fTRIMP, incorporating a continuous approach without prerequisite lactate threshold testing.

7.1.3 Strengths and weakness of TRIMP algorithms

Table 7.3: TRIMP Algorithms by Derivation

		Summation Approach		
		Average heart rate	Discrete sum *(heart rate zones)*	Continuous sum *(exponential curve)*
Weights	Arbitrary		Edwards, Lucia	
	Group/Literature	Banister	Stagno, fTRIMP	Gonzalez-Fimbres
	Individual			iTRIMP

Although the formulas differ, the various TRIMP algorithms are all interconnected, with two of the defining features being the derivation of the weighting factors and the method of summation over time. As illustrated in Table 7.3, heart rate weightings can be based either on arbitrary values, established or group-derived approximations of BLvHR response, or an individual's own BLvHR response curve. Session summary scores can be calculated using either a singular average heart rate value, a summation over time in various heart rate zones, or a summation across individually weighted observations. All of the aforementioned TRIMP algorithms are well-defined and differentiated by these two features. The only missing distinction in the organization of table 7.3 is the difference between Edwards' TRIMP and its arbitrary zone thresholds and the physiologically based thresholds of Lucia's TRIMP. The unlabeled areas in the bottom row of the chart also represent the possibility for other methods of TRIMP calculation. For example, one could use an individualized BLvHR response curve to weight average heart rate for a steady-state session or develop individualized weights and zones for a discrete summation approach.

The analysis conducted demonstrates both the strengths and limitations of the various TRIMP algorithms, with iTRIMP being the preferred approach. In terms of the weighting factors, it is clearly evident that using physiologically based values rather than arbitrary integers increases the validity of the internal training load measure. Similarly, using individualized versus established or group-based weightings represents an improvement, as physiological demands differ across individuals exercising at the same heart rate reserve value (Manzi *et al.*, 2009; Manzi *et al.*, 2013; Malone and Collins, 2016). Finally, considering the summation approach, in the same way that an integral is more accurate than a Reimann sum for calculating area under a curve, continuous summation better represents the physiological demands at all points during a session (Gonzalez-Fimbres *et al.*, 2019).

It is also worthwhile to note a mathematical distinction in established continuous and physiologically based discrete algorithms, with a multiplicative factor of HRR incorporated in the continuous but not the discrete algorithms. In the physiologically based discrete algorithms (sTRIMP and fTRIMP), the weighted value for a given heart rate is determined by approximated blood lactate for that heart rate zone (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019). Therefore, heart rate is used to approximate blood lactate, with the sum of lactate scores over the session determining TRIMP, but heart rate itself is not directly factored into TRIMP. On the other hand, in the continuous algorithms, blood lactate is approximated for a given HRR value via the exponential term, and then multiplied by HRR (Manzi *et al.*, 2009; Gonzalez-Fimbres *et al.*, 2019). As a result of the HRR term in the continuous algorithms (HRR in Equations 7.1 and 7.4), both HRR and extrapolated blood lactate are factors in the weightings. Therefore, in order to provide a discrete approximation of the continuous algorithms, discrete algorithms should incorporate a heart rate factor in their weightings. This would be calculated by taking the mean heart rate reserve value for a given zone and multiplying it by the weighting factor determined by blood lactate. Although the overall form of the algorithm would not change, the individual weights would be altered, with a larger difference for lower heart rate values. As a result, the current discrete algorithms of sTRIMP and fTRIMP, both of which did not incorporate a heart rate weighting term in their development (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019), are overvaluing exercise at lower intensities relative to the continuous models. The units for TRIMP are arbitrary, so a linear transformation of scores would have little impact if the same algorithms were used consistently; however, the multiplicative heart rate term has a non-linear impact on TRIMP weightings. Future physiologically based discrete algorithms could be established with a multiplicative heart rate term used during development of the weightings to align with the continuous algorithms.

With Manzi's iTRIMP combining an individualized and continuous summation approach, it is often considered to be the best TRIMP measure, with this result supported in the literature (Manzi *et al.*, 2009; Akubat *et al.*, 2012; Manzi *et al.*, 2013; Sanders *et al.*, 2017; Fox *et al.*, 2018; Malone *et al.*, 2018). Compared with other internal training load measures, a systematic review of training load in team-sport athletes found iTRIMP to have the strongest correlations (large to very large) with athlete fitness change overtime, the primary metric for validation of TRIMP algorithms (Fox *et al.*, 2018). However, despite iTRIMP being the preferred measure, there are still many situations in which other TRIMP measures may be more appropriate. The biggest barrier to the use of iTRIMP is the requirement for all individuals to complete a lactate threshold assessment prior to monitoring. As lactate threshold testing requires time and resources not always available, group-based TRIMP algorithm such as Gonzalez-Fimbres' TRIMP may be preferable (Gonzalez-Fimbres *et al.*, 2019). Additionally, not all heart-rate monitoring software has the capacity to calculate TRIMP using a continuous summation. For example, Catapult's Openfield Software (Catapult Sports, Version 2.5.0, Melbourne, Australia) allows users to customize the heart rate zones and weights for calculating discrete TRIMP, but not input constants for a continuous approach. Therefore, in this instance, a discrete summation approach would be preferable, such a Stagno's TRIMP, fTRIMP, or even an individually derived discrete algorithm. Finally, the mode of exercise is also important to consider, with Banister's TRIMP and its use of average heart rate being easier to calculate and more accurate for steady state activities.

In conclusion, TRIMP algorithms have evolved over time beginning with Banister's TRIMP and progressing to Manzi's iTRIMP. With its individualized weightings and continuous summation approach, iTRIMP has the strongest physiological basis. Therefore, when resources allow, iTRIMP is currently the best method for monitoring internal training load via heart rate in intermittent ball sports.

7.2 Developing a new testing protocol for calculating iTRIMP in intermittent field-sport athletes

7.2.1 Limitations of the current testing protocol

Although Manzi's iTRIMP is considered to be the best heart rate-based measure of internal training load (Manzi *et al.*, 2009; Manzi *et al.*, 2013; Sanders *et al.*, 2017; Fox *et al.*, 2018; Malone *et al.*, 2018), it is not without its limitations. At the foundation of all TRIMP algorithms is the concept of weighted heart rate values, summated or otherwise adjusted for session duration (Banister, 1991). The strength of iTRIMP over other TRIMP algorithms is that heart rate weights are not only based on BLvHR curves, but also that the curve used is individualized (Manzi *et al.*, 2009; Fox *et al.*, 2018). However, the BLvHR curve for iTRIMP is only as good as the protocol used to derive it. Even at equivalent intensities, the blood lactate response has been shown to vary based on the type of exercise (Akubat and Abt, 2011; Jean-Christophe *et al.*, 2018). It follows that if the testing protocol from which the BLvHR curve is derived does not mimic the demands of the activity iTRIMP is being used to monitor, the accuracy of summary scores is reduced. Thus, a large limitation of iTRIMP lies in the fact that the treadmill test from which BLvHR curves are derived in no way mirrors the intermittent nature of the team-sport activities which iTRIMP is frequently used to monitor (Akubat and Abt, 2011; Fox *et al.*, 2018; Jean-Christophe *et al.*, 2018). Specifically, the testing protocol lacks ecological validity because it is laboratory-based, uses a motorized treadmill, the running in each stage is continuous, and there is no change of direction.

Laboratory-based testing protocols lack ecological validity, with treadmills impacting the physiological demands of running due to the movement of the motorized belt (Jones and Doust, 1996; Padulo *et al.*, 2013; Van Hooren *et al.*, 2020). Although this type of testing allows researchers to control the conditions, the environment in a laboratory differs from that of the field where athletes complete their usual training and competition (Jones and Doust, 1996;

Padulo *et al.*, 2013; Van Hooren *et al.*, 2020). Most notably, the physiological demands of treadmill running differ from overground running. The lack of air resistance when running on a treadmill has been shown to reduce energy cost and oxygen intake compared to non-treadmill running (Jones and Doust, 1996). To adjust for this, many treadmill-based testing protocols, including those for iTRIMP use a treadmill gradient of 1% (Manzi *et al.*, 2013). Although, a treadmill gradient of 1% has been shown to best mirror the energetic demands of outdoor running, it may decrease the validity of the BLvHR curve (Jones and Doust, 1996). Padulo *et al*. reported a significant 5% increase ($p < 0.001$) in heart rate when athletes ran at a treadmill gradient of 2% vs 0% for 5 minutes at a constant velocity due to the increased metabolic demands (Padulo *et al.*, 2013). Blood lactate was also measured at the end of each interval with average blood lactate increasing by 35.5% for the 2% gradient compared to 0% (Padulo *et al.*, 2013). Therefore, although the gradient used in the iTRIMP protocol was only 1%, the data on 2% gradients clearly indicate that even relatively small treadmill gradients impact both heart rate and blood lactate. Furthermore, there is general biomechanical comparability between treadmill and overground running, but a notable difference occurs in the surface stiffness (Van Hooren *et al.*, 2020). As intermittent sports are often played on specialized surfaces (grass, rubber-based turf, water-based turf) of varying stiffnesses, the ecological validity is further reduced, with the impact of this on the BLvHR relationship unclear.

Treadmill-based testing protocols also lack change of direction (COD) which is a common occurrence in field-based sports. Research on the impact of COD on blood lactate is unclear but suggests an increase in blood lactate following shuttle running versus straight-line running (Dellal *et al.*, 2010; Bekraoui *et al.*, 2020). A study comparing the demands of continuous in-line running with shuttle running at 60%, 70%, and 80% of maximal aerobic speed (MAS) found a small but nonsignificant increase in blood lactate and a large, significant increase in heart rate associated with COD (Bekraoui *et al.*, 2020). This supports the earlier work of Dellal *et al.* (2010) who considered the impact of COD in intermittent running with 1:1 work to rest ratios of 10, 15, and 30 seconds at paces of 100-120% MAS. In all instances, blood lactate was significantly higher ($p < 0.01$) following running performed in a shuttle fashion (2 - 3 COD per repetition) compared to straight-line running (Dellal *et al.*, 2010). An increased heart rate was associated with COD for all time durations, but this difference was only significant for the 30 second intervals (Dellal *et al.*, 2010). Although not all findings were significant, both studies had a relatively small sample size of 10 athletes, and taken together, they would suggest that COD impacts both exercising heart rate and blood lactate, likely due to the increased muscular demands of the deceleration and acceleration required (Dellal *et al.*, 2010; Bekraoui *et al.*, 2020). No research has specifically investigated the impact of COD on the BLvHR relationship, so it is unclear if the increase in blood lactate and heart rate associated with COD is proportional. However, there is sufficient evidence to suggest that COD impacts both heart rate and blood lactate individually and thereby may affect the BLvHR response. As a result, for iTRIMP to appropriately summarize the demands of intermittent sports involving COD, COD should be incorporated into the testing protocol.

There are many practical benefits to field-based testing compared to laboratory-based tests. As portable handheld lactate monitors have become more readily accessible, lactate threshold testing can be easily performed in non-laboratory environments. Field-based testing is more readily accessible to sports teams that do not have laboratory or treadmill access, as teams can use their normal training environment for testing. Also, the number of treadmills is not a limiting factor when testing an entire team. This increase in accessibility would allow more sports teams to use iTRIMP and perform more regular re-testing as BLvHR curves change overtime, further increasing data accuracy (Taylor *et al.*, 2021). Finally, athletes may also be more comfortable on their normal sports pitch than in a laboratory environment, removing some of the psychological impact of testing. Beyond any psychological benefit to the athletes, this also improves the validity of BLvHR response data as pre-performance anxiety may impact athlete heart rate (Oliveira-Silva *et al.*, 2018).

In addition to the use of a treadmill, a further limitation of the current iTRIMP testing protocol is the continuous nature of each of the running stages. One of the advantages of current TRIMP algorithms over Banister's TRIMP is that heart rate is not based on an average session value. For this reason, iTRIMP is often used to monitor intermittent sport athletes, as it allows periods of varying intensities to appropriately contribute to training load scores (Malone *et al.*, 2018). However, the testing protocol for calculating iTRIMP involves 4-minute stages of continuous running at a constant intensity. Akubat and Abt investigated the impact of intermittent running on blood lactate response, specifically in reference to the TRIMP testing protocol (Akubat and Abt, 2011). Twelve athletes completed a 4 x 4-minute TRIMP testing treadmill protocol twice, one time running at a constant speed for each 4-minute level and the other alternating between 15s of faster running and 15s of slower running, with the same average pace (Akubat and Abt, 2011). The results showed that the BLvHR relationship is altered by intermittent exercise particularly at higher intensities (Akubat and Abt, 2011). When used to derive TRIMP weightings according to the iTRIMP protocol, the weightings based on the intermittent protocol for HRRs of 0.9 and 1.0 were significantly higher ($p < 0.05$) with a large effect size (1.4 - 1.7) (Akubat and Abt, 2011). Thus, when a continuous testing protocol is used to determine iTRIMP for intermittent exercise, iTRIMP disproportionately underestimates the physiological load at higher intensities. This underestimation not only reduces the validity of iTRIMP values for intermittent activity but may also put athletes training at high intensities at risk for overtraining and overuse injuries.

7.2.2 A field-based, intermittent protocol for calculating iTRIMP in hockey athletes

A new testing protocol was designed to replace the treadmill lactate threshold test used to calculate BLvHR response curves for iTRIMP. It maintains the same overall framework of 4-minute levels with a 1-minute rest but is intended for use on an outdoor sports pitch. To complete the test, athletes run in a rectangle with paces dictated by an audio file. The size of the rectangle, and thus the paces, increases for each level. Regardless of the level, athletes are to complete each side of the rectangle in 10s, with the different side lengths causing athletes to alternate between a slower and faster running speed within each level. Every minute athletes complete a 180° COD instead of the usual 90° COD to run back along the same side they just completed. The pitch setup is shown in Figure 7.1 with the paces and distances shown in Table 7.4.

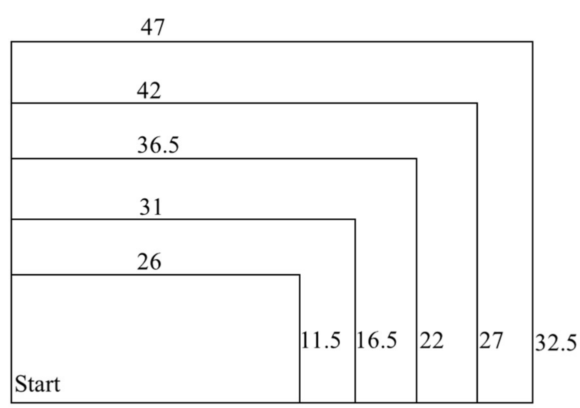


Figure 7.1 Pitch based testing protocol field setup

Table 7.4: Pitch based testing protocol paces and distances

Level	Pace 1 (km·hr^{-1})	Dist. 1 (m)	Pace 2 (km·hr^{-1})	Dist. 2 (m)
1	4.25	11.5	9.75	26.0
2	6.25	16.5	11.75	31.0
3	8.25	22.0	13.75	36.5
4	10.25	27.0	15.75	42.0
5	12.25	32.5	17.75	47.0

Testing protocol:

1. Athletes complete four or five 4-minute levels with a 1-minute rest between level during which blood lactate is measured. Athletes whose blood lactate is greater than 4.0 mmol·L^{-1} after the fourth level are not required to complete the final level.
2. Within each level the athlete is to run around the outside of the box, starting with the long side, taking 10s to complete each side of the box.
3. Pace is dictated by an audio file that beeps every 10s to alert the participant when they should be at a corner of the box. There is also a halfway alert provided after 5s to help athletes guide their pacing.
4. Athletes are instructed to stay as close as possible to the pace dictated by the audio file so that they are reaching the corners of the box in accordance with the beep. If an athlete arrives at a corner too early, he or she should wait rather than commencing the next side early. Verbal encouragement should be provided when required to help athletes maintain the correct pace.
5. Every minute (1.5 laps of the box), the athlete should complete a 180° turn rather than a 90° turn and run back along the same side of the box that they just completed. Athletes should be provided reminders of this either through the audio file or verbally.
6. The first level commences in the smallest box with each subsequent level in a larger box. It is recommended that the various boxes be demarcated with different color cones (for example: blue, green, yellow, orange, red) to prevent confusion.

7.2.3 Test development

The aim of the proposed testing protocol was to determine a BLvHR curve that mirrors the BLvHR response during hockey. This curve can then be used to determine accurate iTRIMP values for summarizing the demands of hockey training and competition. Although the best way to evaluate the physiological response to hockey would be monitoring hockey directly, it is not feasible to regularly monitor blood lactate during hockey. Additionally, the unstructured nature of hockey would make monitoring impossible to control across different individuals and dates. Therefore, this testing protocol was designed to mirror the overall structure of established protocols for calculating TRIMP, with several key changes to the design of the individual levels to increase the ecological validity and sport-specificity.

7.2.3.1 Overall test structure

On a macro level, the overall structure of the testing protocol matches that of existing protocols for calculating TRIMP and specifically iTRIMP (Manzi *et al.*, 2013). The 4-minute levels have been maintained, with heart rate taken as mean heart rate during the last minute of each level and a 1-minute rest between levels during which blood lactate is measured. The 4-minute levels allow sufficient time for athletes to reach a relative steady state at each level, while also not extending the testing protocol unnecessarily, allowing for completion in 20-25 minutes. Additionally, as in the iTRIMP protocol, athletes complete either four or five levels of the assessment dependent on when blood lactate exceeds 4 $mmol \cdot L^{-1}$, representing the onset of blood lactate accumulation (Manzi *et al.*, 2009). Although athletes often complete a ramp to exhaustion upon completion of the submaximal portion of the treadmill testing protocol, this was not incorporated here because the aim of this test was to determine BLvHR curves, not evaluate athlete fitness levels nor establish athletes' maximum heart rate. However, it is important to note that accurate resting and maximal heart rate values are required for calculation of heart rate reserve. Therefore, if athletes' maximum heart rates are not known, then a maximal fitness test, such as the 30-15 intermittent fitness test, could be performed (Buchheit, 2010).

7.2.3.2 Change of direction

The rectangular layout of the test allows for the incorporation of both 90° and 180° COD. Field-based sports such as hockey which are stochastic in nature (McGuinness *et al.*, 2017; McGuinness *et al.*, 2019) require athletes to change direction at all angles. However, it is not possible to incorporate all angles into a testing battery. Therefore, 90° and 180° COD were chosen as the two angles to be included in the protocol as 180° COD represents the most extreme change athletes would regularly complete and the 90° CODs requires athletes to pivot in both directions. The incorporation of the 180° turn within the testing protocol, not only allows for the incorporation of this turning angle, but also ensures that athletes are regularly changing the direction of the 90° turns so as not to be completing all 90° turns to either the right or left. Additionally, as a result of the 180° turns, athletes run at the same pace in two consecutive ten second intervals three times during each level. Although this reduces the frequency of pace changes, it also decreases the consistency and repetitiveness of the testing protocol (albeit in a controlled and repeatable fashion), better representing the unpredictable nature of team sports.

From a practical perspective, a box is relatively easy to set up, with two sides of the box aligned with sideline/baseline markings already on most sports pitches. However, one drawback of this design is that the rectangular shape increases the pitch-space requirements compared to a straight-line shuttle, reducing the number of athletes that can complete the testing protocol simultaneously. On a standard hockey or football pitch, four athletes could complete testing simultaneously, with one setup on each half of the pitch and two athletes starting in opposite corners of each rectangle. As blood lactate needs to be measured at the end of each stage, the number of sports scientists and portable lactate analyzers is more likely to be a constraint for teams rather than pitch space.

The frequency of COD within the testing protocol is based on an analysis of COD within hockey competition. As part of the recovery monitoring study in chapter 5, positioning and movement data was collected from seventeen female athletes (21.3 ± 2.4 years, 167 ± 4 cm, 62 ± 5 kg) across four hockey matches. Not all athletes competed in all matches, and, in total, 47 match files were included in the analysis. Matches were part of England Hockey's Northern Division One league, the second highest level of domestic competition in England. During competition athletes wore Catapult S7 10 Hz GNSS units (Catapult Vector S7, Catapult Sports, Melbourne, Australia) which contain a triaxial accelerometer collecting data at a frequency of 100 Hz. As part of its inertial movement analysis (IMA), Catapult reports COD and accelerations/decelerations grouped by the direction of the movement. For the purpose of this analysis, COD was grouped into two categories: right/left COD (-135°, -45°) or (45°, 135°), and total COD (-135°, 135°). In hockey competition, right/left COD occurred on average 4.77 times per minute and total COD occurred 0.77 times per minute. Thus, the testing protocol was designed in accordance with these frequencies with each 4-minute stage containing twenty 90° COD (5 per minute) and three 180° COD (0.75 per minute). As a result, the COD contained within the testing protocol matches both the frequency and approximates the direction (albeit in a more controlled fashion) of the COD in hockey competition.

7.2.3.3 Intermittent running

The testing protocol was designed to better simulate the physiological demands of team sports such as hockey by incorporating intermittent speed changes rather than continuous running within each level. As intermittent running can take many forms, there were decisions made regarding the number of paces used, the length of the intervals, and the distances/paces. In all cases, the aim was to balance sport-specificity with validity and accessibility of the testing protocol.

The first consideration for intermittent running was the choice to alternate between two speeds per level. The alternating pattern was based on the work of Akubat and Abt, who first suggested using an intermittent testing protocol for the calculation of iTRIMP and demonstrated the significant impact of an intermittent protocol on iTRIMP weighting values, particularly at high intensities (Akubat and Abt, 2011). As with the COD, incorporating only two speeds per level is a vast simplification of the running patterns in hockey. However, this is where practicality becomes important, with the choice to use only two running paces per level based on accessibility of the testing protocol. Requiring athletes to alternate between three or more running speeds, particularly in a field-based test where pace is controlled by the athlete not by belt speed on a treadmill, increases the difficulty of the test from a pacing perspective. This would have elevated the risk of athlete error and would have increased the learning effect, meaning athletes would need more familiarization with the protocol prior to testing.

A secondary consideration of intermittent running was the length of the individual intervals, with the framework for interval length following the work of Akubat and Abt, who used alternating fifteen second intervals (Akubat and Abt, 2011). However, the intervals were shortened in this protocol to mirror the average effort length in hockey. Using the same 47 match files as for COD calculations, an analysis was performed using Catapult Openfield software (Catapult Sports, Version 2.5.0, Melbourne, Australia) to determine the average effort length in hockey competition. To match the minimum speed incorporated in the testing protocol, as illustrated in Table 7.3, the velocity threshold was set to 4.25 km·hr^{-1}. Thus, effort length was defined as the distance covered by an athlete in an individual instance where they exceeded 4.25 km·hr^{-1} until their speed dropped below that threshold. For the 47 match files analyzed, average effort length was determined to be 26.0 m. The average paces of the testing protocol were predetermined to match those of the original treadmill assessment, with a mean pace of 11.0 km·hr^{-1} for those athletes who completed all five stages and 10.0 km·hr^{-1} for those who stop after the fourth stage. Taking an average speed of 10.0 or 11.0 km·hr^{-1}, a 26.0 m interval would last either 9.36s or 8.51s seconds. Therefore, a shorter effort length of approximately 8.94 seconds would best match the effort lengths in hockey competition. The match data for average effort length and change of direction frequency linked well with average effort lengths of 8.9s and change of direction frequency of 5.54 instances per minute (once per 10.8s). Thus, to simplify the testing protocol for participants, minimizing the opportunity for error and decreasing the learning effect, the decision was made to coincide changes of speed with CODs. As athletes naturally decelerate to change direction, the corners make an obvious choice for where to also change speed. Therefore, an interval length of 10s was chosen, approximating both the change of direction frequency and effort length in hockey competition.

The shorter interval length also positively impacted athlete pacing as the ten second intervals permit more regular feedback to the athlete on their pacing, with the halfway alert coming every five seconds. Regular pacing feedback is particularly important for this testing protocol as athletes are setting their own speed, using the audio cues to guide them. Although five second feedback would be possible within fifteen second intervals, visualizing a third of a distance is not as intuitive, so more cones would have been required, further complicating the setup. Thus, the shorter intervals not only make the protocol more sport-specific, but they also increase accessibility and promote compliance with the prescribed pacing.

7.2.3.4 Running speeds and distances

The average running speeds were selected based on the running paces used in the treadmill based TRIMP protocols, with the variation determined by pace distributions in hockey competition. The distances used for each level were calculated so that athletes ran, on average, the same speeds as during the treadmill version of the iTRIMP protocol. These speeds were adjusted from Manzi's original levels to be appropriate for the female athletes in the study and to match the speeds used previously in female university hockey athletes of 7, 9, 11, 13, and 15 $km \cdot hr^{-1}$ (Konerth, 2019).

To make the running intermittent in nature, athletes alternated between running 2.75 $km \cdot hr^{-1}$ above and below the average speed per level. This variation of 2.75 $km \cdot hr^{-1}$ was derived from the variation in speed distributions during 20s intervals of hockey competition. To determine this value, an analysis was again performed using the 47 hockey match files from the observational study on recovery monitoring (Chapter 5). However, instead of performing the analysis in Catapult Openfield, raw files were exported into Microsoft Excel (Microsoft Corporation, Version 2002, Redmond, Washington) for evaluation. To minimize the quantity of computations required, a stratified random sample of match data was used, with 5 minutes analyzed per match file. Each of the four matches was randomly assigned one of the four quarters and a random number generator was then used to determine the precise start time for the five-minute analysis period. The 10 Hz velocity data was separated into 20s intervals for analysis. Note that 20s intervals were used instead of 10s intervals as athletes' speeds were designed to stay constant within 10s intervals but instead vary over two 10s periods in the testing protocol. Visual inspection of velocity traces was used to determine when athletes made substitutions, with any 20s intervals during which an athlete was not on the pitch for the duration of the 20s excluded from the analysis. Histograms of athlete speed showed speed was positively skewed, so a non-parametric approach was taken with interquartile range considered in favor of standard deviation. The interquartile range of velocity was calculated for each 20s period, with the average value across intervals determined to be 5.5 $km \cdot hr^{-1}$. Thus, in alignment with the variation of velocity within 20s intervals in hockey competition, running speeds were taken as 2.75 $km \cdot hr^{-1}$ above and below the average velocity for each level.

When calculating distances based on the prescribed paces for each interval, adjustments were made to account for COD. COD takes time, with turns at various angles shown to increase sprint times compared to straight-line running (Buchheit *et al.*, 2010; Buchheit, Haydar and Ahmaidi, 2012). As a result, if COD is not accounted for in interval distances, athletes will run faster than the target speed during each interval. Similar adjustments are made in other shuttle-based running protocols (Bangsbo, 1994; Buchheit, 2008). For example, when developing the 30-15 intermittent fitness test, Martin Buchheit used a time adjustment of 0.7 seconds per 180° COD when calculating distances. However, as 90° COD is rarely incorporated into fitness assessments, there is no standard adjustment. Over a 25 m sprint, team sport athletes were shown to have a time-increase of 30.0% when there was a 180° COD compared to straight-line running (Buchheit *et al.*, 2010). Similarly, over a 30 m sprint with two 90° COD, time was shown to increase by 34.4%, or 17.2% per COD (Buchheit, Haydar and Ahmaidi, 2012). Considering the percentage decrement in sprint times and Buchheit's adjustment of 0.7s, an approximation for 90° COD can be calculated as follows.

$$17.2\%/30\% \times 0.70s = 0.40s$$

As every sixth turn in the testing protocol is a 180° COD, an average adjustment for COD in the assessment was calculated as

$$5/6 \times 0.40 + 1/6 \times 0.70 = 0.45s.$$

Therefore, when determining the length of each interval, distance was taken as that covered when running for 9.55s at the target speed rather than 10s. This adjustment is limited for several reason. Specifically, it is based on percentage decrement over maximal sprints rather than moderate running, the 90° COD was based on two turns in one shuttle, and the shuttles were different lengths. Additionally, a different adjustment was not used for the 90° and 180° CODs as the interval distances and times were kept the same. However, as this assessment was designed to determine the BLvHR relationship rather than assess velocity at set blood lactate values, these limitations were accepted and considered to not notably impact any results.

7.2.4 TRIMP calculation

Individualized TRIMP can be calculated from this testing protocol using the same approach as traditional iTRIMP (Manzi *et al.*, 2009). In short, an individual's BLvHR reserve is plotted for each stage, with an exponential curve approximated via least squares regression. This curve is of the form $y = ae^{bx}$ with individualized constants a and b, and iTRIMP is calculated via equation 7.3. In addition, this protocol can be used to develop team TRIMP protocols for a group of athletes, if individual TRIMP calculations are not possible. To develop a continuous team TRIMP algorithm, the same procedure would be used as for iTRIMP, except the groups' combined BLvHR reserve values would be used in place of an individual's values. To determine a discrete team TRIMP algorithm, the group BLvHR curve would be calculated and heart rate zones determined with blood lactate approximated for the average value in each heart rate zone. As discussed in section 1.3, this value would then be multiplied by the mean HRR for the zone to determine weighting values. Thus, this submaximal lactate threshold testing protocol can be used to determine an appropriate TRIMP algorithm regardless of the calculation approach required.

7.2.5 Conclusion

Overall, the aim of this testing protocol was to determine a BLvHR curve for an individual that best represents the relationship between heart rate and blood lactate in hockey competition. As the BLvHR curve is what determines the iTRIMP algorithm, the accuracy of iTRIMP is only as good as the testing protocol used. To increase the ecological validity of the testing protocol, it was adapted from a laboratory-based design to a field-based test that athletes could complete on their normal playing surface. The length of the levels and average speed of the levels were taken from the laboratory-based protocol to maintain the established structure and allow for comparison. However, rather than continuous running within each level, speeds alternated between 10s of faster and slower running, with the 10s interval length based on the average effort length in hockey competition. The paces of the faster and slower intervals differed by 5.5 $km \cdot hr^{-1}$, the average interquartile range of velocity over 20s intervals in hockey competition, and 90° and 180° CODs were incorporated in alignment with the frequency of these CODs in hockey. This protocol also increases the accessibility of iTRIMP monitoring to team-sport programs who do not have access to laboratory testing.

Chapter 8: A Comparison of pitch and laboratory-based individualized training impulses over a hockey season

8.1 Introduction

Based on an athlete's physiological response to exercise, Manzi's iTRIMP is often considered to be the gold standard TRIMP measure (Manzi *et al.*, 2009; Sanders *et al.*, 2017; Fox *et al.*, 2018). However, as intermittent running changes the blood lactate versus heart rate (BLvHR) relationship, iTRIMP may misrepresent the physiological load of athletes (Akubat and Abt, 2011). A range of pitch-based testing protocols have been established to evaluate other fitness markers in field-sport athletes (Bangsbo, Iaia and Krustrup, 2008; Buchheit, 2010; Shushan *et al.*, 2022). For example, the Yo-Yo intermittent recovery test and 30-15 intermittent fitness test are commonly used maximal tests to approximate maximal oxygen consumption (Bangsbo, Iaia and Krustrup, 2008; Buchheit *et al.*, 2009; Buchheit, 2010). Additionally, a range of field-based submaximal testing protocols have been developed and evaluated in intermittent-ball sport athletes (Shushan *et al.*, 2022). These testing protocols not only address the issue of ecological validity but also increase the accessibility of testing to large groups of athletes, with pitch-based fitness tests such as the Yo-Yo and 30-15 widely adopted and frequently used both in the literature and in applied settings (Bangsbo, Iaia and Krustrup, 2008; Buchheit, 2010). However, to date, the only established protocol for determining BLvHR curves for individualized TRIMP algorithms is laboratory-based. Although, TRIMP testing differs from the above protocols in that blood lactate measurements are required, the accessibility of handheld portable lactate analyzers mitigates this problem. Therefore, a pitch-based submaximal lactate threshold test for determining iTRIMP algorithms would not only better mirror the demands of intermittent-sport activity but also increase the accessibility of iTRIMP monitoring.

Manzi's iTRIMP algorithm is structured similarly to the original TRIMP algorithm proposed by Banister, which includes the product of an exponential weighting term and exercising heart rate (Banister, 1991). The exponential weighting term is based on an individual's BLvHR curve with the output of that term approximating blood lactate for the inputted heart rate. However, unlike discrete TRIMP models such as Stagno's TRIMP and fTRIMP (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019), continuous models such as

Gonzalez-Fimbres and Manzi's TRIMP, also directly include heart rate reserve as a separate term in the TRIMP algorithm (Manzi *et al.*, 2009; Gonzalez-Fimbres *et al.*, 2019). As HRR is already incorporated into the exponential term, multiplying by HRR is, at best, redundant and, at worse, skewing TRIMP scores, with time spent at lower heart rates contributing less to overall TRIMP. Therefore, this study evaluated the effectiveness of this term in continuous TRIMP algorithms by comparing TRIMP loads calculated with and without this term.

Effective training load monitoring can improve performance, reduce injuries and overtraining, and elevate athlete fitness levels (Casamichana, Castellano and Castagna, 2012; Kevin and James, 2015; Mara *et al.*, 2015; Bourdon, 2017). However, athlete monitoring is only as effective as the monitoring technique used. TRIMP measurement in field sport athletes is currently limited in that the fitness test from which the algorithms are derived does not mirror the demands experienced by athletes, and the heart rate reserve term in continuous TRIMP equations may skew load scores. Therefore, this study aimed to evaluate and compare individualized TRIMP calculated via a new pitch-based testing protocol (piTRIMP) with Manzi's iTRIMP, and iTRIMP and piTRIMP calculated both with and without the heart rate reserve term, over the course of a hockey season. TRIMPs were analyzed via a comparison of TRIMPs with other training load metrics and a consideration of the dose-response relationship between cumulative TRIMPs and fitness changes. These analyses were used to determine if there was a significant difference between TRIMP algorithms, and, if so, which algorithm(s) best predicted fitness change.

8.2 Methods

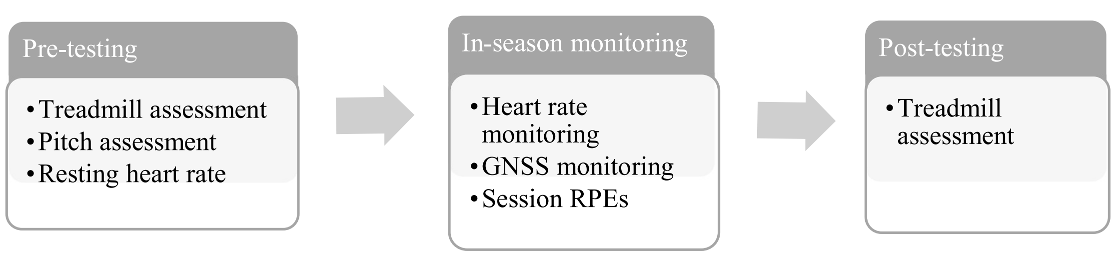


Figure 8.1 Training Impulse Study Design

An observational approach and repeated measures design were utilized with the study taking place for a period of nine weeks during a competitive hockey season. The study design is outlined in Figure 8.1. Athletes began by completing both a treadmill-based and pitch-based submaximal lactate threshold test. From these tests, two BLvHR curves were extrapolated for each athlete and used to determine individualized algorithms for both pitch-based TRIMP and Manzi's iTRIMP, with and without the HRR term. Velocity at 4 $mmol \cdot L^{-1}$ during the laboratory-based assessment was taken as the onset of blood lactate accumulation and used as a marker of athlete fitness (Manzi *et al.*, 2009). Athletes' internal training loads (session rating of perceived exertion and heart rate) and external training loads (total distance and high speed running) were monitored during bi-weekly trainings and competitions for a period of nine weeks during the hockey season. At the conclusion of the study, athletes repeated the laboratory-based fitness test to allow for an analysis of the dose-response relationship between training load metrics and athlete fitness.

<u>8.2.1 Participants</u>

This study began with 17 female participants from Durham University Hockey Club's first team (21.4 ± 1.9 yrs, 165.5 ± 5.2 cm, 63.5 ± 6.4 kg). Participants had been playing hockey for 12.5 years on average and had completed a five-week training block to prepare for the start of the season immediately prior to data collection. Durham University Hockey Club's women's first team competes in England Hockey's National League and British University Colleges Sport National League. As the hockey club's first team match-day squad consists of 15 outfield athletes, with some movement due to form and availability, this sample size was chosen to allow for monitoring of all athletes competing at the first team level at the start of the season. Goalkeepers were excluded due to the significant difference in the demands of their position. Prior to the start of the study, all participants completed a prescreening questionnaire to ensure that they were free from serious injury and were not at an elevated risk of cardiovascular complications from exercise (Appendix D). Ethical approval was obtained from the university ethics board, and all participants were required to provide informed consent (Appendix G). All relevant government, university, and England Hockey guidelines were followed in relation to COVID-19.

Although 17 athletes began the study, 10 athletes were included in the final analyses (22.0 ± 2.1 yrs, 165.3 ± 5.2 cm, 61.4 ± 5.3kg, 13.1 ± 3.2 yrs experience). To be included, athletes were required to have performed post-testing and have heart rate and GNSS data for at least 80% of training and competition sessions. Due to injuries, frequent changes in athlete selection and non-compliance with wearing monitors, seven athletes did not reach this minimum threshold of 80% and were excluded from the analysis. Although the rate of exclusion was very high, this was likely due to the study being performed in the first full hockey season following COVID-19. Athletes had not played together during the previous season, so there was increased uncertainty around athlete selection and athlete movement between teams. Additionally, due to an equipment malfunction during the fitness testing and another athlete not completing the testing, a final sample size of nine athletes was used for the evaluation of iTRIMP and eight for piTRIMP.

8.2.2 Procedures

8.2.2.1 *Resting and maximal heart rate*

To assess resting heart rates, athletes were instructed to lie quietly, without distraction, for five minutes upon waking while wearing a heart rate monitor (Polar T10, Polar Electro, Kempele, Finland). The minimum value recorded was taken as an athlete's resting heart rate. Athletes' maximal heart rate was taken as the highest heart rate value recorded either during hockey competition, training, or maximal testing. For maximal testing, most participants in the study had completed a 30-15 intermittent fitness test within the past 12 months (Buchheit, 2010). Participants who had not completed the 30-15 fitness test performed a ramp to exhaustion immediately following the lactate threshold treadmill test with speed increasing 0.5 $km \cdot hr^{-1}$ every 30s until volitional exhaustion to establish maximum heart rate.

8.2.2.2 *Lactate threshold testing*

Athletes completed three fitness testing sessions, the pitch-based protocol at the start and the laboratory-based protocol at both the start and end of the study. Testing was arranged around the athletes' training schedule to minimize fatigue, and participants were advised to avoid strenuous activity and abstain from alcohol consumption 24 hours prior to testing. When possible, individual's testing was scheduled at a similar time of day to minimize the impact of circadian variation. Due to laboratory and pitch availability, all athletes completed the treadmill testing protocol first. All testing was completed at the Durham University Sport and Wellbeing Park, either in the physiology laboratory or on the hockey pitches.

The laboratory lactate threshold test was based on the iTRIMP protocol developed by Manzi *et al.* (2009). The test was performed on a treadmill (H/P/Cosmos Pulsar, H/P/Cosmos Sports and Medical GmbH, Germany) and commenced at a speed of 7 $km \cdot hr^{-1}$, increasing by 2 $km \cdot hr^{-1}$ each stage up to a maximum of 15 $km \cdot hr^{-1}$. Treadmill gradient was set at 1% to mimic the energetic costs of outdoor running (Jones and Doust, 1996). Stages were 4 minutes with a 1-minute recovery between stages during which fingertip blood lactate measurements were taken using a handheld lactate analyzer (Lactate Plus, Nova Biomedical, Waltham, WA, USA). To account for varying fitness levels in the group and to ensure testing was submaximal, athletes were only required to complete stages up to a blood lactate accumulation of at least 4 $mmol \cdot L^{-1}$. Throughout testing, athletes wore a heart rate monitor (Polar H1, Polar Electro Oy, Kempele, Finland) and a Catapult GNSS device (Catapult Vector S7, Catapult Sports, Melbourne, Australia). The heart rate monitor was worn on a strap around the chest and the GNSS monitor between the shoulder blades in the pouch of a specially formulated vest. Although GNSS data were not used during testing, the Catapult devices were used to record and download heart rate data, consistent with the procedure for in-season monitoring.

Athletes also completed a pitch-based lactate threshold assessment, as outlined in Chapter 7. This assessment was performed on a hockey pitch to ensure consistency with the athletes' normal playing surface and environment. Athletes wore the same heart rate and GNSS monitors, and fingertip blood lactate was again analyzed using a handheld lactate analyzer. Athletes were encouraged to stay at the appropriate speeds as dictated by the beeps and were audibly reminded of the 180° turns every 60 seconds.

8.2.2.3 *Training load measurement*

Training load measurement occurred during participants' regularly scheduled training and competition for the duration of the study. Although there was some variation in weekly schedule due to match dates and coaching decisions, a typical week consisted of training sessions on Mondays 20:00-22:00 and Fridays 7:15-8:45, with matches on Wednesday and Saturday afternoons.

Athlete monitoring was performed via heart rate and GNSS monitors. External training load was measured using Catapult's Vector S7 GNSS devices (Catapult Vector S7, Catapult Sports, Melbourne, Australia), which were previously tested and shown to be a valid and reliable measure of hockey movement patterns (Chapter 6). External training load was measured in terms of total distance (TD) and high speed running (HSR), which was defined as > 15 $km \cdot hr^{-1}$. The 15 $km \cdot hr^{-1}$ threshold has been proposed for use in hockey and is in alignment with previous research (Jennings *et al.*, 2012a; Jennings *et al.*, 2012c; Ihsan *et al.*, 2017; Sunderland and Edwards, 2017; Polglaze *et al.*, 2018; Morencos *et al.*, 2019). Internal training load was measured using heart rate monitors (Polar T10, Polar Electro, Kempele, Finland), with both iTRIMP and piTRIMP calculated from heart rate data. To provide complete data on athlete load, distance covered during warm-up and cool-down periods was included in the analyses. Athletes wore the same heart rate and GNSS monitors for the duration of the study.

Rating of perceived exertion (RPE) was also collected for each training session and competition using a combination of Foster's adaption and the Borg 100-point scale (Table 8.1) (Foster *et al.*, 2001; Borg and Borg, 2002). This scale was chosen in favor of a 10-point or 14-point scale due to its intuitive nature and to allow for more precise responses, and Foster's adaption of categories and a fixed 0-100 scale was utilized as it has been shown to be effective for the calculation of session RPEs (Foster *et al.*, 2001). The scale and anchors were explained to the athletes and athletes were asked to provide an overall response for the full training session. To minimize the impact of the Hawthorne effect wherein participants change their behavior as a result of being monitored (Buckworth, 2002), athletes were clearly instructed that RPEs would not be shared with coaches and that there were no 'correct' responses. RPEs were collected via an online google form, with athletes asked not to discuss their answers with others to reduce the influence of peer pressure. Athletes were prompted to fill out the form following sessions, with a reminder sent out to those who had not submitted. Responses were then used to calculate sRPE, the product of session duration and RPE, with times determined from GNSS data (Foster *et al.*, 2001).

Table 8.1: Rating of Perceived Exertion Scale (Foster et al., 2001).

RATING	DESCRIPTOR
0	Rest
10	Very, Very, Easy
20	Easy
30	Moderate
40	Somewhat Hard
50	Hard
60	.
70	Very Hard
80	.
90	.
100	Maximal

8.2.3 Analysis

8.2.3.1 *TRIMP algorithms*

Based on the data recorded during fitness testing, iTRIMP was calculated for each athlete following the procedures outlined by Manzi *et al.* (2009) . Exercising heart rate for each stage was taken as average heart rate in the final minute of each level and plotted against blood lactate measurements recorded in the rest period immediately after. BLvHR curves were smoothed, and exponential curves were fitted via least squares regression. These curves were then used to determine the two unique constants for each individual, *a* and *b*.

Equation 8.1 Blood lactate versus heart rate reserve

$$Blood\ Lactate = ae^{b(HRR)}$$

Equation 8.2: Heart rate reserve

$$Heart\ Rate\ Reserve\ (HRR) = \frac{Exercising\ HR - Resting\ HR}{Maximum\ HR - Resting\ HR}$$

Individualized TRIMP algorithms were then calculated as follows, where *a* and *b* were unique constants for each individual and *n* was the number of heart rate readings per minute.

Equation 8.3: iTRIMP1 and piTRIMP1

$$iTRIMP1/piTRIMP1 = \frac{1}{n}\sum_{HR}\left(\frac{HR - HR_{rest}}{HR_{max} - HR_{rest}}\right) \times ae^{b\left(\frac{HR-HR_{rest}}{HR_{max}-HR_{rest}}\right)}$$

Note the 1 is used to distinguish this original method of calculation from the second method of calculation without the heart rate reserve term outlined below. This procedure was performed twice for each individual using the data from the laboratory-based and pitch-based lactate threshold testing, with training load scores from the laboratory-based algorithms denoted as iTRIMP1 and from the pitch-based algorithm denoted as piTRIMP1. Additionally, to evaluate the impact of the heart rate reserve term, modified TRIMPs were also calculated without this term for both iTRIMP and piTRIMP. These are denoted as iTRIMP2 and piTRIMP2 and were calculated as follows with a and b being the same unique constants for each individual and testing procedure (pitch or treadmill).

Equation 8.4: iTRIMP2 and piTRIMP2

$$iTRIMP2/piTRIMP2 = \frac{1}{n}\sum_{HR} ae^{b\left(\frac{HR-HR_{rest}}{HR_{max}-HR_{rest}}\right)}$$

The terms piTRIMPs and iTRIMPs will be used to denote both pitch-based (piTRIMP1 and piTRIMP2) or both laboratory-based (iTRIMP1 and iTRIMP2) metrics, respectively.

8.2.3.2 *Athlete fitness*

The dose-response relationship between training load and fitness changes over the course of the study was evaluated to assess which training load markers best predicted fitness changes. For alignment with the literature and as the purpose of this study was not to evaluate the pitch-based protocol as a measure of athlete fitness (although this could be evaluated in future research), athlete fitness was assessed via the laboratory-based lactate threshold test, rather than the pitch-based assessment. Velocity during submaximal testing has been shown to be more sensitive to fitness change than maximal assessments (Impellizzeri, Rampinini and Marcora, 2005). Additionally, athlete motivation plays a smaller role than in maximal tests such as the 30-15 IFT, where testing concludes when athletes reach volitional exhaustion. Therefore, a blood lactate concentration of 4 mmol·L^{-1} was taken as the onset of blood lactate accumulation, with velocity at this value (v_{OBLA}) taken as a marker of athlete fitness. To calculate v_{OBLA}, athletes' blood lactate at the completion of each stage was plotted against velocity. Via exponential interpolation, velocity at 4 mmol·L^{-1} (v_{OBLA}) was then extrapolated for each individual. Changes in v_{OBLA} over the course of the study were taken as measures of fitness increase or decline, with higher velocities indicating elevated fitness.

8.2.3.3 *Data analysis*

Prior to statistical analysis, data was downloaded, visually inspected, exported and processed. Data were downloaded from the Catapult S7 devices using Catapult's OpenField software (Catapult Sports, Version 2.5.0, Melbourne, Australia). For all sessions, heart rate and velocity were plotted for each individual and visually inspected for erroneous data. Erroneous data were due to user error or equipment malfunction and were clearly evident, with participants either late to put on their monitoring equipment or with heart rate intermittently dropping to zero and spiking. Erroneous data were flag and removed from the analyses, with those segments marked as missing data.

Heart rate and GNSS data were exported to Microsoft Excel (Microsoft Corporation, Version 2211, Redmond, Washington) for further analysis. Summary values for GNSS data (total distance and HSR) were exported directly from Openfield. However, Openfield was unable to calculate continuous TRIMPs, so heart rate data was exported as raw files of individual athletes' recorded heart rate for each session, measured at a frequency of 10 Hz. Beginning with the code published by Mitch Henderson for tidying Openfield raw files (Henderson, 2020), adding in TRIMP functions referencing individuals' unique values, and iterating over a working directory of raw files organized by training session, R (R Core Team, 2022) was used to calculate iTRIMP1, iTRIMP2, piTRIMP1 and piTRIMP2 for each individual for each training session. Heart rate and GNSS data were then aligned, along with sRPEs, to produce an overall training load dataset. Missing data were approximated via training/competition sessions of the same type, individually weighted based on a comparison with typical group mean.

8.2.3.4 *Statistical analysis*

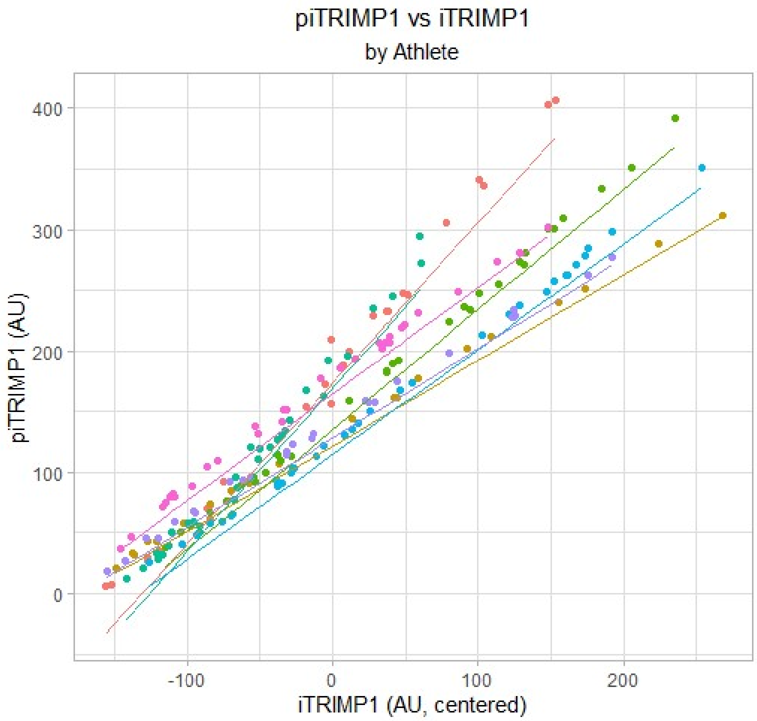
Data were checked for normality and outliers using visual inspection of histograms and stem and leaf plots. Although iTRIMPs and piTRIMPs were found to be somewhat right skewed, the regression techniques used for the analyses were considered to be sufficiently robust to avoid the need for transformation. The first analysis was a comparison and modeling of piTRIMPs vs iTRIMPs over the course of the season. Although it is still common practice to ignore the interdependence of repeated measures data in sport research, this practice can often result in skewed or invalid results (Bland and Altman, 1995). Therefore, to account for the lack of independence caused by repeated measurements on individual athletes, a multilevel modeling was performed. The model was built up from the empty model with Akaike's information criterion (AIC) checked as additional effects were added. A random slope and random intercept model was used to model piTRIMPs based on iTRIMPs for each athlete. Session type (training or match) was also investigated as a predictor but was not found to improve model fit. Due to the relatively small sample size, modeling was performed using restricted maximum likelihood to minimize the bias of variance estimates. Grand mean centering was performed to provide meaningful intercept values and allow for comparisons of equivalent training load scores across individuals (Enders and Tofighi, 2007). Models are presented alongside 95% confidence interval estimates for fixed effects and the standard deviation of random effects. Analyses were performed in R (R Core Team, 2022) with the nlme package v3.1-159 (Pinheiro *et al.*, 2021), and graphs produced via ggplot2 v3.3.6 (Wickham, 2016).

The next analysis was a comparison of TRIMPs with other training load measures. The repeated measures for individual athletes created dependence and made a standard regression analysis inappropriate (Bland and Altman, 1995). Repeated measures correlations were selected as an investigatory analysis revealed that allowing slopes to vary did not greatly improve model fit compared to a random intercept model. Additionally, unlike with piTRIMPs vs iTRIMPs, the primary outcome of interest was correlation rather than intercept and slope values (Bland and Altman, 1995). This distinction was because the purpose of the analysis was not to investigate any of these training load metrics being used in place of iTRIMP, as would be the case with piTRIMP, but rather to compare the interrelatedness of the various metrics with both iTRIMP and piTRIMP. Repeated measures correlations were calculated using the Rmcorr package v0.5.2 (Bakdash and Marusich, 2017). Due to the limitations of traditional null hypothesis significance test p-values (Wasserstein, Schirm and Lazar, 2019), the minimum effects tests was used to calculate p-values for the correlation coefficients of the repeated measures correlations (Lakens, Scheel and Isager, 2018). Given that the metrics tested were all training load measures, one would reasonably expect a clear correlation between values (McLaren *et al.*, 2018). Therefore, traditional p-values for the correlation coefficient indicating a significant dependence between variables would be of little practical relevance. Instead, a test was performed to determine if variables were highly correlated, as defined as $|r| > 0.7$, with (-0.7, 0.7) taken as the range of no practical significance. Laken's spreadsheet for equivalence testing (Lakens, 2017) was used to perform two one-sided tests, with the p-value for the minimum effects tests being the inverse (1-x) of the result of the equivalence testing (Lakens, Scheel and Isager, 2018).

The final analysis performed was an investigation of the dose-response relationship between athlete training load and fitness changes over the course of the season. As training load was considered in terms of total season load with only one measure per athlete, observations were independent, and a traditional least squares linear regression was performed with Pearson product moment coefficients reported. In alignment with the previous analysis, p-values for correlation coefficients are expressed as the result of minimum effects testing. However, a dose-response relationship between training load and fitness change was not assumed, so a test was performed to determine the existence of a substantive correlation, with trivial correlation coefficients of (-0.1, 0.1) taken as the range of no practical significance. Regression data are presented alongside 95% confidence intervals, with statistical significance set to $p < 0.05$.

8.3 Results

There was a total of 322 sessions analyzed, of which 195 were trainings and 127 were matches. The relationships between iTRIMPs and piTRIMPs are visualized in Figure 8.2, with various colors and regression lines representing different athletes. iTRIMP1 significantly predicted piTRIMP1, $b = 0.97$, $t(190) = 9.88$, $p < 0.001$. The relationship between iTRIMP1 and piTRIMP1 showed significant variance in intercepts across athletes, $SD = 27.3$ (95% CI: 15.5, 48.1), $p < 0.001$. In addition, slopes varied across athletes, $SD = 0.261$ (95% CI: 0.148, 0.462), $p < 0.001$, with slopes and intercepts significantly positively correlated cor = 0.837 (95% CI: 0.384, 0.965), $p = 0.005$. Similarly, iTRIMP2 significantly predicted piTRIMP2, $b = 0.98$, $t(190) = 10.0$, $p < 0.001$. Both intercepts and slopes varied significantly across athletes, $SD = 36.8$ (95% CI: 20.8, 64.9) $p < 0.001$ and $SD = 0.259$ (95% CI: 0.146, 0.457) $p < 0.001$, respectively. Slopes and intercepts were positively correlated, but this result was not significant, cor = 0.527 (95% CI: -0.215, 0.883), $p = 0.145$.



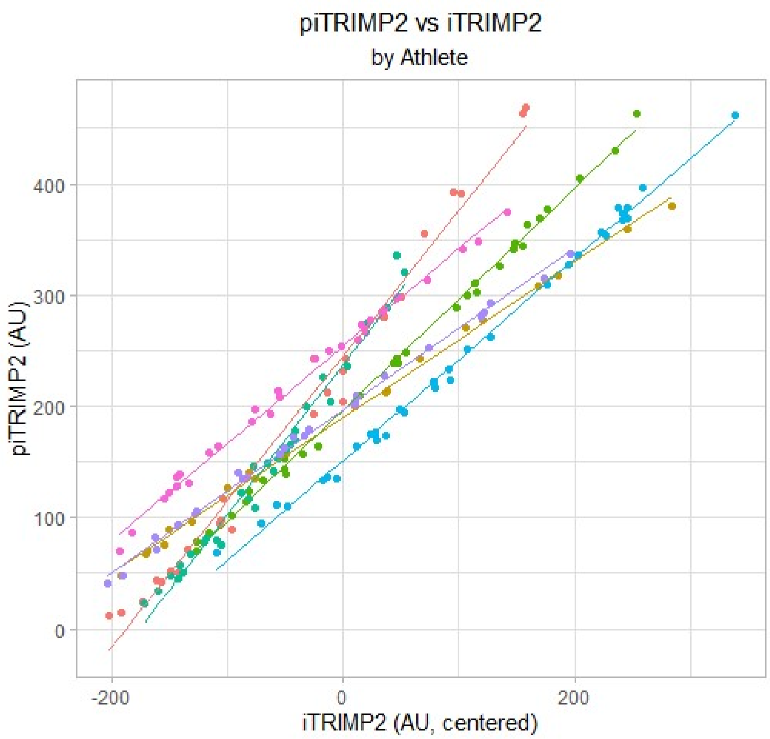


Figure 8.2   piTRIMP vs iTRIMP correlations

Table 8.2: Correlation of TRIMP with other training load metrics

	iTRIMP1			iTRIMP2		
	r	95% CI	p^	r	95% CI	p^
sRPE	0.766	(0.703, 0.817)	0.020	0.772	(0.709, 0.821)	0.012
Total Distance	0.930	(0.909, 0.946)	<0.001	0.934	(0.914, 0.949)	<0.001
High Speed Running	0.789	(0.733, 0.835)	0.002	0.791	(0.735, 0.836)	0.001
	piTRIMP1			**piTRIMP2**		
	r	95% CI	p^	r	95% CI	p^
sRPE	0.794	(0.731, 0.839)	0.001	0.801	(0.745, 0.845)	<0.001
Total Distance	0.911	(0.885, 0.931)	<0.001	0.920	(0.897, 0.938)	<0.001
High Speed Running	0.774	(0.715, 0.822)	0.008	0.778	(0.720, 0.825)	0.005

sRPE: session rating of perceived exertion. r: pearson product moment coefficient. CI: confidence interval. p^: significance calculated via a minimum effects test for a strong correlation |r|>0.7

The repeated measures correlations between TRIMPs and total distance, HSR, and sRPE are summarized in Table 8.2. The strongest correlations were between TRIMPs and total distance. In all instances, correlations were slightly stronger between iTRIMP2/piTRIMP2 and other training load metrics than with iTRIMP1/piTRIMP1; however, these differences were small and not significant. Similarly, there were no significant differences in the strength of the correlations between piTRIMP1 and iTRIMP1 or between piTRIMP2 and iTRIMP2. The minimum effects tests evaluating the likelihood of a strong correlation (|r| > 0.07) between TRIMPs and other training load metrics were significant in all cases ($p < 0.05$).

Table 8.3: Dose-Response Relationship of TRIMPs with Fitness Change

	r	95% CI	p^
iTRIMP1	0.856	(0.445, 0.969)	0.002
iTRIMP2	0.940	(0.734, 0.988)	<0.001
piTRIMP1	0.687	(-0.035, 0.938)	0.049
piTRIMP2	0.914	(0.588, 0.985)	0.001

r: pearson product moment coefficient. CI: confidence interval. p^: significance calculated via a minimum effects test for a nontrivial correlation |r|>0.1

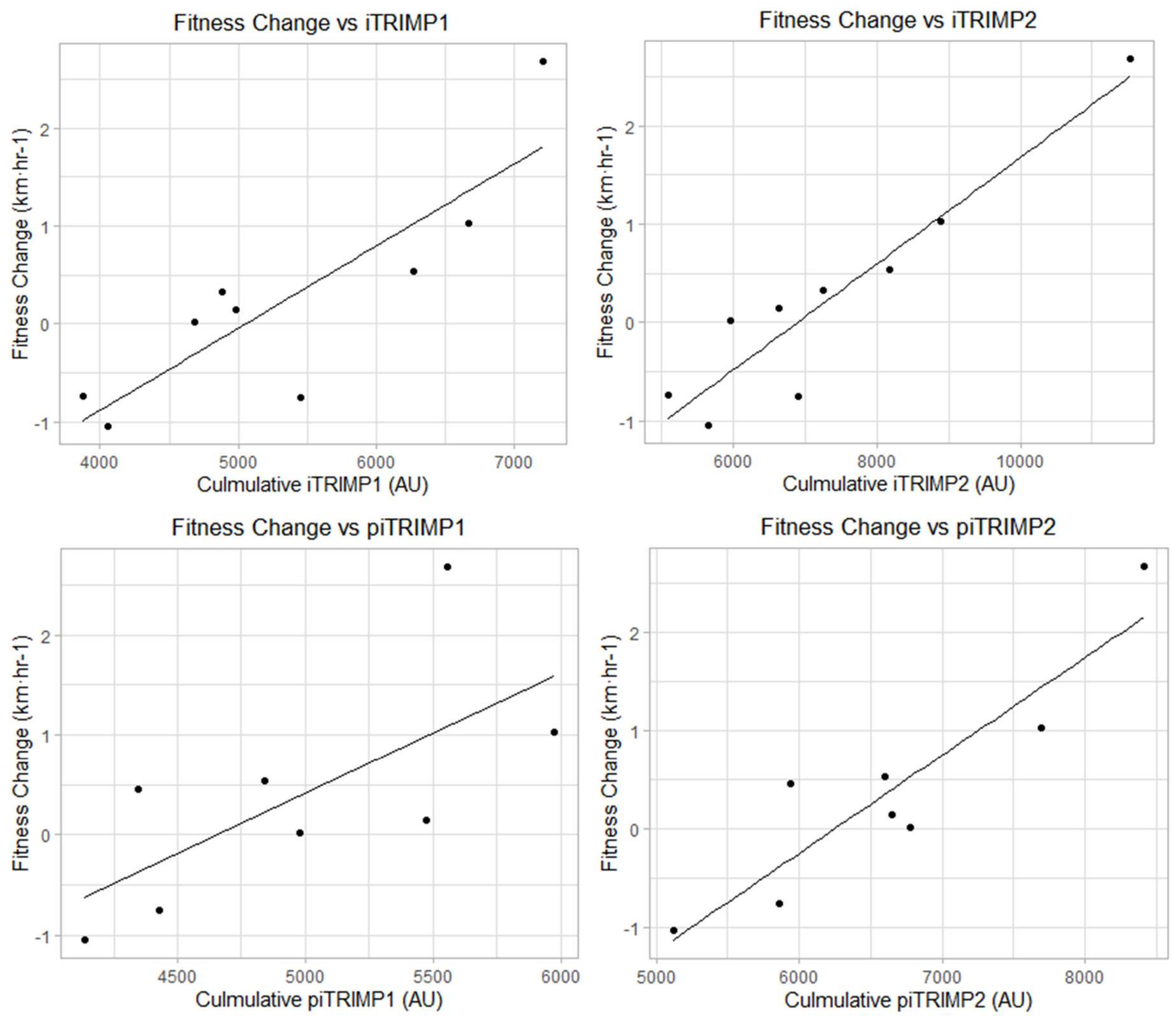


Figure 8.3: Dose-Response Relationship of TRIMPs with Fitness Change

All TRIMP metrics had a significantly non-trivial dose-response relationship with athlete fitness change over the course of the season (Figure 8.3 and Table 8.3). The correlations between fitness change and iTRIMP2/piTRIMP2 (r = 0.940, 0.914) were stronger than those between fitness change and iTRIMP1/piTRIMP1 (r = 0.856, 0.914), but these differences were not significant. Similarly, the correlations with iTRIMP1 and iTRIMP2 were stronger, but not significantly so, than piTRIMP1 and piTRIMP2. One athlete had a notably greater fitness improvement than the other athletes. This value was statistically not an outlier, so it was included in the analysis. However, due to the small sample size and potential for this value to skew the correlations, the analysis was also performed without this value, with no significant differences found in the results.

8.4 Discussion

The results of this study indicate that piTRIMP is a unique method of calculating internal training load, with scores distinct from the laboratory-based metric. Thus, piTRIMPs should not be used interchangeably with iTRIMPs across athletes. Despite this, both pitch-based and laboratory-based metrics were comparably strongly correlated with other internal (sRPE) and external (TD and HSR) training load metrics. TRIMPs calculated using the established algorithm with the heart rate reserve term (iTRIMP1 and piTRIMP1) showed a weaker dose-response relationship with fitness change than when this term was omitted (iTRIMP2 and piTRIMP2). Therefore, piTRIMP2 and iTRIMP2 would appear to be the best training load metrics for use in hockey athletes. Due to its increased ecological validity and accessibility, piTRIMP may be the preferable metric; however, future studies will be required to evaluate the differences between piTRIMP2 and iTRIMP2.

8.4.1 Pitch-based versus laboratory-based iTRIMPs

Although strongly correlated, the relationships between both piTRIMP1 and iTRIMP1, and piTRIMP2 and iTRIMP2 varied significantly between athletes, indicating that piTRIMPs and iTRIMPs were distinct metrics. In other words, the pitch-based testing protocol was not an outdoor replication of the laboratory-based protocol, and the two measures should not be used interchangeably across athletes. Given their similar algorithm derivations, iTRIMPs and piTRIMPs were strongly correlated, with iTRIMPs significantly predicting piTRIMPs within individuals ($p < 0.001$). This may appear to suggest that load scores derived via piTRIMP were just a linear transformation of iTRIMP scores, and not distinct measures of internal load, especially given the arbitrary nature of TRIMP units. However, the relationships between piTRIMPs and iTRIMPs differed significantly between individuals. Firstly, there was significant variance in intercepts, with a standard deviation of intercepts of 27.3 AU (p/iTRIMP1) and 36.8 AU (p/iTRIMP2). As a grand mean centering technique was used, this indicates that for a hypothetical session with consistent iTRIMP loads across athletes (in this case 1:169.5 AU and 2: 233.5 AU) the corresponding piTRIMP loads for that same session would vary with a standard deviation of 27.3 and 36.8 AU between athletes. Thus, a session with an identical internal load measured via iTRIMP would be recorded as significantly harder or easier for some athletes when monitored via piTRIMP. Change of direction and acceleration/deceleration, both of which are included in the piTRIMP protocol but not iTRIMP, are skills that can be trained, with efficiency and technical proficiency varying across athletes (Cronin and Hansen, 2006; Nygaard Falch, Guldteig Rædergård and van den Tillaar, 2019; Cormier *et al.*, 2020). Therefore, it follows that the BLvHR relationship may respond differently in various athletes when these skills are added into the testing protocol. As intermittent sports incorporate change of direction, change of speed, and acceleration/deceleration at similar frequencies as the piTRIMP testing protocol, the internal load scores derived from piTRIMP may be more reflective of the true demands of these sports than iTRIMP, but more research will be needed to confirm this hypothesis.

In addition to the difference in intercepts across athletes for piTRIMPs vs iTRIMPs, it is also worthwhile to note the significant variance in slopes (1: $SD = 0.261$, $p < 0.001$; 2: SD = 0.259, $p < 0.001$) and the positive correlation between intercepts and slopes (1: cor = 0.837, p = 0.005; 2: cor = 0.527, p = 0.145). The variance in slopes indicates that not only do some athletes have higher/lower load scores when piTRIMPs are used instead of iTRIMPs (variance in intercepts), but also that for sessions of varying intensities the relationships between piTRIMPs and iTRIMPs change between athletes. Furthermore, the positive correlation between intercept and slopes (significant for p/iTRIMP1 but not for p/iTRIMP2) demonstrates an increasing discrepancy between piTRIMPs and iTRIMPs between athletes as iTRIMPs increase. In other words, as session intensity, as measured by iTRIMPs, increases, there is a greater discrepancy in piTRIMP scores across athletes. When session load is higher, this is usually due to more time being spent in higher heart rate zones, which have a larger weighting value in TRIMP calculations (Manzi *et al.*, 2009). Therefore, the discrepancy between iTRIMPs and piTRIMPs is likely the result of blood lactate accumulating more rapidly in some athletes than others at elevated heart rates when intermittent running and change of direction are incorporated. This finding is in alignment with the work of Akubat and Abt who investigated the difference in blood lactate concentrations and derived iTRIMP weightings when using a continuous versus intermittent running protocol during testing (Akubat and Abt, 2011). Akubat and Abt reported that iTRIMP weightings (and thus blood lactate accumulations) were significantly higher in athletes completing an intermittent rather than a continuous running protocol at HRR values of 0.9 (continuous: 7.04 ± 0.72 AU, intermittent: 8.07 ± 1.73 AU) and 1.0 (c: 9.20 ± 1.22 AU, i: 11.25 ± 2.65 AU) (Akubat and Abt, 2011). In addition, at these higher intensities the standard deviations of weighting scores were also more than doubled in the intermittent versus the continuous testing protocol (0.72 vs 1.73 AU and 1.22 vs 2.65 AU), in agreement with the increasing discrepancy across athletes at higher intensities when running is intermittent (Akubat and Abt, 2011).

The difference between iTRIMPs and piTRIMPs, particularly for more intense sessions, can largely impact athlete training load scores, which has implications for athlete fitness, overtraining, and injuries. As session intensity rises, it is increasingly important for training load scores to be an accurate reflection of athlete demands because higher load scores have a greater impact on athlete cumulative load (Andrade *et al.*, 2020; Maupin *et al.*, 2020). However, it is precisely for these sessions that piTRIMPs and iTRIMPs differ most, indicating, in alignment with the work of Akubat and Abt, that the demands of high intensity intermittent exercise may be underestimated by iTRIMPs (Akubat and Abt, 2011). The results of this study go further to suggest that this underestimation may be athlete-dependent with some athletes impacted significantly more than others. Two systematic reviews have reported strong associations between acute:chronic workload ratios (ACWR) and risk of injury (Andrade *et al.*, 2020; Maupin *et al.*, 2020), with ACWR also a good indicator for safe return to play post injury (Blanch and Gabbett, 2015). Specifically, maintaining an ACWR of 0.8-1.3 has been shown to reduce injury risk and can be used to improve fitness and performance (Blanch and Gabbett, 2015; Gabbett, 2016; Andrade *et al.*, 2020; Maupin *et al.*, 2020). Given the relative specificity of these ratios, and large impact of the most recent training sessions on ACWR, when using the recommended exponentially weighted approach (Murray *et al.*, 2017), it is clear the importance of accurate training load scores for athlete monitoring. Therefore, it is critical to note the uniqueness of these training load metrics and ensure that an accurate and consistent metric is used across athletes and over a monitoring period. Given the strong within-athlete correlations of piTRIMPs and iTRIMPs, it is possible that piTRIMP could be calculated as a simple linear transformation of iTRIMP or vice-versa for an individual athlete. However, as this transformation would be unique to the athlete, it would be impossible to determine without the athlete completing both testing protocols, at which point it would be of little practical value. Additionally, as the intermittent ball sports for which this protocol were designed are team sports, athlete monitoring, even if individualized, is likely being performed across groups of athletes rather than for individuals.

8.4.2 Relationships with other load markers

Both iTRIMPs and piTRIMPs showed comparable, significantly strong correlations with other internal and external training load metrics. The correlation between TRIMPs and external training load are stronger than those previously reported in a meta-analysis of team sport athletes. Specifically, the meta-analysis reported a pooled effect for TRIMP versus total distance of 0.74 and TRIMP versus high-speed running of 0.28 (McLaren *et al.*, 2018). However, the correlation with total distance was based on only two studies, one in soccer and one in Australian rules football, both of which have different physiological demands than hockey (Scott *et al.*, 2012; Scott *et al.*, 2013). TRIMP was also calculated using Banister's and Edward's algorithms, with the limitations of these algorithms (use of average heart rate, generic zones and weightings, not individualized) likely contributing to the decreased correlations (Scott *et al.*, 2012; Scott *et al.*, 2013). Similarly TRIMP versus HSR correlations were based on seven studies across a range of sports including rugby, soccer, Australian football and basketball, with TRIMP calculated using various algorithms and only one study using iTRIMP (McLaren *et al.*, 2018). However, previous research into university hockey reported similar associations as in this study with a correlation between iTRIMP and total distance of r = 0.882 and between iTRIMP and high speed running (15-19 $km \cdot hr^{-1}$) of r = 0.707 (Konerth, 2019). Despite being another internal training load measure, the correlations with sRPE were not stronger than those with the external load markers. However, sRPE was calculated with total session time, rather than just active time, potentially reducing the specificity of this metric (Konerth, 2019).

The strength of the correlations with other training load measures provides good evidence in support of the use of the individualized TRIMPs for monitoring hockey athletes. As internal and external training load are distinct constructs, a perfect correlation would not be expected (Impellizzeri, Marcora and Coutts, 2022). However, as physiological demands are related to physical outputs, it has been suggested that internal versus external load associations can be used to assess the construct validity of internal training load measures (McLaren *et al.*, 2018). Therefore, the very strong association of the TRIMP metrics with total distance and strong association with high-speed running provides evidence in support of these as valid markers of internal training load. There were no significant differences between iTRIMP1 and iTRIMP2 and between piTRIMP1 and piTRIMP2 indicating that the heart rate reserve term in the algorithm does not impact the associations with other training load metrics. Despite the differences between iTRIMPs and piTRIMPs, the correlations of iTRIMPs and piTRIMPs with other load markers were not significantly different. This may be due in part to the use of a repeated measures correlation instead of a multilevel model, despite preliminary analyses indicating only small improvements in model fit with random slopes.

8.4.3 Dose-response relationship with fitness changes

Unlike external training load where metrics such as distance and velocity are clearly defined, measures of internal training load are often harder to qualify (Impellizzeri, Marcora and Coutts, 2022). There is no criterion value for the physiological response to exercise so internal load metrics are often evaluated based on their dose-response relationship with fitness change (Thomas, 2011). Thus, the results of this study would indicate that iTRIMP2 and piTRIMP2 are valid internal training load metrics and may be better suited for monitoring intermittent sport athletes than iTRIMP1 and piTRIMP1. Both iTRIMP2 ($r = 0.940$) and piTRIMP2 ($r = 0.914$) had notably stronger correlations with athlete fitness change than iTRIMP1 ($r = 0.856$) and piTRIMP1 ($r = 0.687$). This finding would suggest that the additional heart rate reserve term in continuous TRIMP algorithms does not improve the modeling of athlete internal load. However, due in part to the small sample sizes, these differences were not significant, so more studies will be needed to evaluate these relationships. When iTRIMP2 and piTRIMP2 are considered, the dose-response relationship with fitness change is comparable. These results indicate that both metrics are good internal load measures and future research will be needed to evaluate differences in the two. Thus, despite being distinct metrics, when used consistently iTRIMP2 and piTRIMP2 are similarly effective at predicting athlete fitness change.

The results of this study demonstrate a stronger association between athlete fitness and iTRIMP than has been previously reported. Multiple studies have investigated the dose-response relationship between iTRIMP and changes in v_{OBLA}, with a correlation of $r = 0.78$ ($p < 0.01$) reported in hurling (Malone and Collins, 2016; Malone *et al.*, 2018) and $r = 0.64$ ($p = 0.004$) in premiership soccer athletes (Manzi *et al.*, 2013). Notably, a much weaker association was found in youth soccer athletes with $r = 0.33$ ($p > 0.05$) for v_{OBLA} but $r = 0.67$ ($p < 0.05$) for velocity at blood lactate levels of 2 $mmol \cdot L^{-1}$ (Akubat *et al.*, 2012). However, this difference may have been due to the age and development of the athletes considered (Akubat *et al.*, 2012). In terms of hockey, an association of $r = 0.597$ was found between iTRIMP and percent fitness changes (measured as the average change in velocity at 2 and 4 $mmol \cdot L^{-1}$) in university hockey athletes (Konerth, 2019).

The training response is influenced by other factors, in addition to internal training load, including sleep, recovery, stress, and nutrition (Sperlich and Holmberg, 2017). Therefore, the fact that 88% and 84% and of the variability in athlete fitness were explained by iTRIMP2 and piTRIMP2 scores, respectively, indicates that these metrics do well to summarize the internal physiological demands of hockey athletes.

8.4.4 Limitations

There are several key limitations of this study that are important to note. Firstly, this study was largely limited by the small sample size both in terms of the number and breadth of athletes included. Future studies should evaluate these relationships with more athletes across teams and sports. Additionally, athletes were only monitored during their on-pitch hockey training with athletes' daily activity levels and any outside training not evaluated as part of this study. Although this is commonplace in studies evaluating training load metrics, this is clearly a limitation as athletes' off-pitch actions impact both on-pitch performance and fitness changes (Sperlich and Holmberg, 2017). Athletes' BLvHR relationship was measured in preseason training and used throughout the season. Again, although this is common with individualized training load monitoring, the BLvHR relationship can change as fitness changes over the course of the season (Fox, Scanlan and Stanton, 2017; Fox *et al.*, 2018). These within-athlete changes are likely to be smaller than between-athlete changes thereby still suggesting the use of individualized monitoring; however, given the pitch-based, submaximal nature of the testing protocol, mid-season testing could be easily introduced if resources allow, and that level of specificity is desired.

Finally, the impact of athletes' menstrual cycles was not considered. Approximately half of the participants reported taking hormonal birth control prior to the start of the study, which decreases hormonal variations over the course of the menstrual cycle (NHS Digital, 2017). Therefore, with the small same size and this further distinction between participants, there was not adequate power to evaluate this factor. Additionally, previous research on university athletes has shown that lactate threshold, as measured during the laboratory-based fitness testing in this study, is not significantly different across stages of the menstrual cycle (Bossi *et al.*, 2013; Ross *et al.*, 2017).

8.5 Practical applications

This study supports the separate use of both iTRIMP2 and piTRIMP2 as valid internal training load metrics for monitoring hockey athletes. Compared with iTRIMP1 and piTRIMP1, iTRIMP2 and piTRIMP2 had stronger dose-response relationships with athlete fitness change, suggesting that the heart-rate reserve term should be removed from continuous TRIMP algorithms. piTRIMP2 and iTRIMP2 had similarly strong relationships both with fitness change and other training load metrics; however, a direct comparison of these metrics indicated that they are distinct metrics and cannot be used interchangeably across athletes. Therefore, the results of this study would suggest that either a laboratory-based or pitch-based protocol can be used to calculate individualized TRIMP in hockey athletes, as long as the same protocol is used across all athletes. This increases the accessibility of individualized internal load monitoring, which improves the accuracy of athlete monitoring. The pitch-based TRIMP protocol may be preferable over the laboratory-based protocol due to its increased ecological validity, but future research will be needed to investigate the relationship between these metrics, particularly across a larger sample of athletes.

Chapter 9: Overall Discussion – An Evidence-Based Model for Athlete Monitoring in Hockey

9.1 Introduction

Innovations in technology have resulted in increased athlete monitoring in hockey and other intermittent ball sports (Torres-Ronda *et al.*, 2022). Primarily focused around training load and recovery, athlete monitoring can help maximize performance by reducing injuries, increasing fitness and improving wellbeing (Bompa, 1999; Gabbett and Domrow, 2007; Stagno, Thatcher and Van Someren, 2007; Cummins *et al.*, 2013; Drew and Finch, 2016; Eckard *et al.*, 2018). However, athlete monitoring is only as good as the technology, protocols, and implementation of the information used (Torres-Ronda *et al.*, 2022). If the monitoring techniques and equipment employed are inaccurate, incorrect conclusions will be drawn due to erroneous information. Similarly, if the protocols and equipment are sound but the variables are not properly evaluated, the technology is of little use. In order to set up an effective athlete monitoring system, various metrics must be appropriately used and their interactions interpreted correctly.

As highlighted throughout this thesis, athlete monitoring is multifaceted and interconnected, with no singular gold-standard measure. For example, internal and external training load measure distinct but interrelated concepts surrounding an athletes' work during a session (Impellizzeri, Marcora and Coutts, 2022). Similarly, athlete recovery is influenced by innumerable factors taking place both on and off the pitch, with sport specific and general life stressors influencing recovery status (Duffield *et al.*, 2018). Although these metrics may be considered in isolation, to fully understand training dose and response, monitoring should be multivariate in its approach (Heidari *et al.*, 2019). Athlete monitoring can quickly become overwhelming, with macro level conclusions obscured by the sheer volume of variables (Torres-Ronda *et al.*, 2022). Therefore, when developing an athlete monitoring system, it is important to be deliberate in variable selection and understand the relationships between the monitoring components (Thornton *et al.*, 2019).

The various components of athlete monitoring in hockey have been critically unpacked across the previous chapters. Internal training load, external training load, and athlete recovery have each been examined in detail to determine the best approaches for measurement. These studies will now be brought together to develop an overall model of athlete monitoring in hockey. Components of the model consist of the aforementioned internal training load, external training load, and recovery monitoring, as well as athlete fitness and performance as key contributing and outcome factors, respectively. This chapter begins with an analysis of the interactions between these components on the macro level. Then, an evidence-based model for athlete monitoring in hockey will be proposed with the individual model components and their validated approaches for their measurement considered in detail based on the results of this research.

9.2 Modeling athlete monitoring in hockey

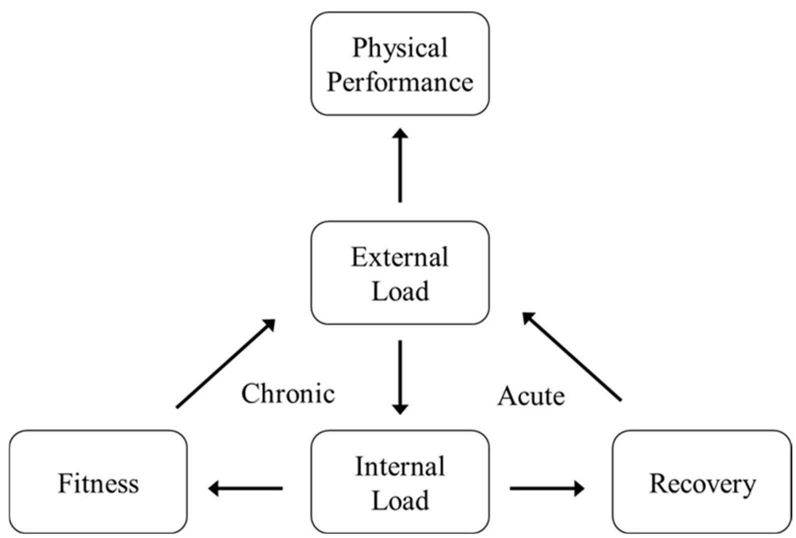


Figure 9.1: Conceptual Model of Athlete Monitoring

As a product of the research for this thesis, a conceptual model for athlete monitoring in hockey was devised and is illustrated in Figure 9.1. It contains the five model components previously discussed, with the arrows representing the relationships between these components at the macro level. This conceptual framework aligns with the recommendations of Impellizzeri *et al.* (2005), Coutts *et al*. (2017), and Gabbett *et al.* (2017), illustrating the various components of the training process. It also incorporates elements of the fitness-fatigue model originally outlined by Banister in 1991. This model builds upon the existing models in the literature (Impellizzeri, Rampinini and Marcora, 2005; Coutts, Crowcroft and Kempton, 2017; Tim *et al.*, 2017) and provides the framework for the evidence-based model presented in the subsequent section. It is important to note that this model is not designed to illustrate every interaction and component that impacts athlete training and performance. There are a myriad of factors and countless interactions that can influence athlete status and physical performance. However, this schematic was designed to provide a conceptual framework of the overall interactions. As such, it can be used by coaches and analysts alike to understand how various athlete monitoring methods are interrelated and can be implemented to improve athletes' physical performance.

The model is centered around external training load, the physical output during a training or competition session. As outlined in Impellizzeri's theoretical framework of the training process and Gabbett's model of the athlete monitoring cycle, external training load dictates the internal training load, or physiological demands on the athletes (Impellizzeri, Rampinini and Marcora, 2005; Gabbett *et al.*, 2017; Impellizzeri, Marcora and Coutts, 2022). Put simply, the amount of work performed impacts the physiological load required to perform the work. This is consistent with the work of Impellizzeri *et al.,* (2005) who noted in their framework that individual characteristics, quality, quantity, and organization impact the external-internal training load relationship. For example, the makeup (quality, quantity, and organization) of the external load matters in terms of how that work is performed. Consider running 1600 m at a set pace versus running 4 x 400 m intervals at the same pace with a 3-minute rest between repetitions. Although the external load is the same in terms of the distance covered and pace, the way that the load was performed will influence internal training load. Thus, external load, the way in which the load is performed, and the individual characteristics of the athlete performing it will determine internal training load (Impellizzeri, Rampinini and Marcora, 2005)

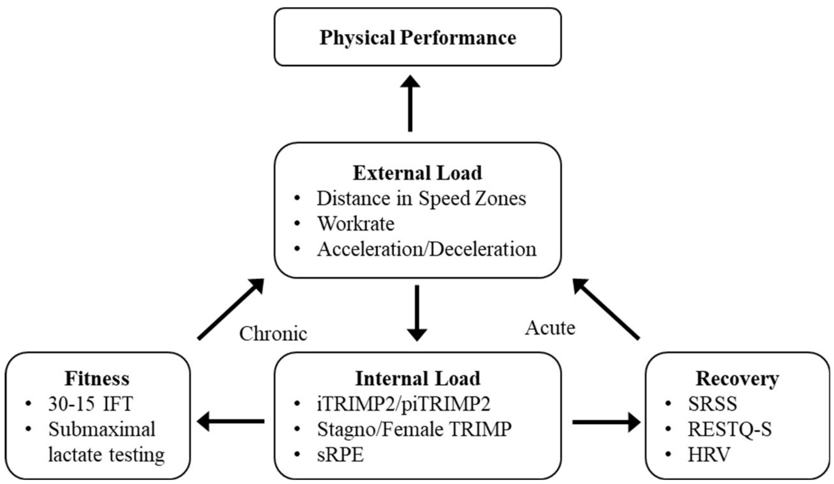
As outlined in the fitness-fatigue model, internal training load contributes to both athlete fitness and fatigue (Banister, 1991). In their chapter on developing athlete monitoring systems, Coutts *et al*. (2017), argued that the fitness-fatigue model should form the theoretical underpinning for athlete monitoring systems. In the newly proposed model presented in this thesis, athlete recovery is used instead of fatigue to better incorporate the entire wellbeing status of the athlete. As an athlete's wellbeing is not solely determined by their workload, overall life stressors and other lifestyle factors are important to consider (Gabbett *et al.*, 2017). Although the nomenclature of recovery in this model is distinct, it is in alignment with the recommendations of Coutts *et al*. (2017) who suggested that subjective measures such as RESTQ-S and SRSS be used to quantify athlete fatigue in monitoring systems. Similarly, Gabbett *et al*. (2017), described this element of the athlete monitoring cycle as perceptual wellbeing, explicitly incorporating external factors that may contribute to athlete recovery status. Thus, recovery is used here to incorporate both the internal training demands and outside factors that together determine an athlete's response to a given stimulus.

Internal training load also directly contributes to athlete fitness (Banister, 1991). Following the principle of progressive overload, when increased training stimuli are implemented, adaptations occur increasing physiological performance (Pearson *et al.*, 2000). As with recovery, the relationship between internal training load and fitness is not closed, with outside factors such as nutrition and sleep also impacting athlete fitness levels. Thus, although not perfectly correlated, cumulative internal training load is predictive of and has a strong dose-response relationship with fitness change over time (Coutts, Crowcroft and Kempton, 2017; Gabbett *et al.*, 2017).

Both athlete recovery and fitness directly contribute to an athletes' readiness to compete and external training load in future sessions (Banister, 1991; Gabbett *et al.*, 2017). As first described by Banister (1991), predicted performance, defined as external training load, can be taken as the mathematical difference of fitness and fatigue (fitness minus fatigue). Thus, this creates a loop, in alignment with the models of Coutts *et al.* (2017) and Gabbett *et al.* (2017), with previous load influencing current training. As fatigue develops and decays faster than fitness, the right side of the model with recovery is more acute, whereas the left side with fitness is more chronic (Banister, 1991; Bosquet *et al.*, 2007; Coutts, Crowcroft and Kempton, 2017). In the case of overtraining, poor recovery can become chronic over time, but this only occurs when acute changes in athlete recovery persist and are not addressed for extended periods (Meeusen *et al.*, 2013). This distinction in the timeline of fitness and recovery loops is important when considering the longer-term development of training programs and is what allows for periodization and taper (Bosquet *et al.*, 2007). Additionally, the timeframes provide insight into the recommended frequency of these aspects of athlete monitoring, with athlete recovery requiring more frequent monitoring than athlete fitness.

Finally, at the top of the model lies physical performance. Physical performance is defined here as the ability to produce work. Thus, external training load, which itself is influenced by recovery and fitness, in large part determines physical performance. However, external training load and physical performance are two distinct constructs, and, as such, are separated in this model. Overall, this newly developed model provides a conceptual framework for the interactions of the various athlete monitoring metrics which can be used to inform decisions when evaluating athlete data and developing athlete monitoring systems.

9.3 Evidence-based model for athlete monitoring in hockey



Recommendations based on the results of chapters 4 & 6 (external load), chapter 5 (recovery), and chapters 7 & 8 (internal load).

Figure 9.2: An Evidence-based Model for Athlete Monitoring in Hockey

Building on the conceptual framework above, Figure 9.2 presents a newly developed evidence-based model for athlete monitoring hockey, incorporating validated monitoring metrics for each component as determined by the results of this thesis. Various metrics are incorporated to allow for consideration of the different aspects of each monitoring component. Additionally, since monitoring techniques differ in the resources required to collect data, multiple measures are included to provide alternative monitoring options. Physical performance as defined here is an abstract concept, not something directly measurable, hence why there is no specific metric delineated. This model illustrates the relationship between the various measures of athlete monitoring and physical performance, and it can be used to inform how athlete monitoring is conducted for hockey athletes. The model components and their subsequent measures will now be critically unpacked based on the results of the individual studies of this thesis.

9.3.1 External training Load

As examined in chapter 4, external training load in hockey is primarily measured via global positioning system (GPS) or global navigation satellite system (GNSS) units (Bourdon, 2017; Malone *et al.*, 2017). Catapult Sports is the primary manufacturer of GNSS units, with Catapult Sports' or a subsidiary's units used in 96% of research into external training load in hockey competition (Table 4.1). Chapter 6 of this thesis showed Catapult Sports' most recent GNSS unit, the Vector S7, to be a valid and reliable measure of athlete movement patterns in hockey via speed and distance. The negative bias of GNSS and GPS units during short change of direction is a notable limitation of these devices which was again highlighted by the validation results of this thesis and indicates an area for future technological improvements. However, these biases are generally considered acceptable (Scott, Scott and Kelly, 2016; Crang *et al.*, 2021). Therefore, Catapult's Vector S7 provides a valid, reliable, and practical measure of external training load in hockey.

In addition to total distance and distance in speed zones, GNSS monitoring systems also provide external load information on a wide range of other variables, with some of the more commonly reported metrics in hockey including max speed, accelerations, decelerations, metabolic power and the proprietary playerload (Table 4.3). With so many variables to consider, it can be hard to understand which external training load measure to focus on and prioritize for monitoring (Crang *et al.*, 2021). The systematic review of hockey competition in chapter 4 showed total distance and distance in speed zones to be the most commonly reported external training load metrics with only 2 out of 28 studies monitoring external training load not reporting at least one of these measures (Vescovi, 2014; Chesher *et al.*, 2019). Similarly, in a meta-analysis of internal and external training load, McLaren *et al. (2018)* reported that 87% of studies included total distance or distance in speed zones as a measure of external training load. As the frequency of measurement would suggest, distance-based measures provide an intuitive and valuable way of quantifying physical load and performance, forming the basis for external training load measurement (Crang *et al.*, 2021).

Several other external load measures in hockey competition were critically unpacked in chapter 5 and could be evaluated as part of an athlete monitoring system. As hockey has rolling substitutions, workrate standardizes distance data, allowing for comparisons across athletes playing different minutes (Konerth, 2019). This metric can be particularly useful as an intensity and physical performance marker; however, in terms of overall load management, overall distances are still important to consider. Accelerations and deceleration in hockey competition were also considered in chapter 4, particularly in terms of comparisons across quarters, between playing positions or at different performance levels (Table 4.3). The results of the systematic review indicated that acceleration and decelerations can also be considered in a performance and fatigue context, with elite hockey athletes having been shown to have increased intensity and frequency of acceleration and decelerations over the course of a match than lower-level athletes (Buglione *et al.*, 2013; Johnston *et al.*, 2015). Additionally, Morencos *et al*. (2018) and Chesher *et al.* (2019) have reported that high-intensity accelerations and decelerations significantly decrease over the course of a hockey match, suggesting that they are a highly sensitive measure of fatigue. Although generally considered secondary in priority compared to distance-based measures, as part of their meta-analysis on the subject, Harper *et al.* (2019) argue for the importance of high-intensity acceleration and deceleration monitoring. Specifically, accelerations and decelerations place distinct metabolic and mechanical demands on athletes and are thus important to consider in terms of load management and injury preventions (Young, Hepner and Robbins, 2012; Gastin *et al.*, 2019; Harper, Carling and Kiely, 2019).

The results of the systematic review and meta-analysis performed on hockey competition clearly indicate that external training load in hockey varies significantly across populations, levels, sexes, and positions. This finding further reinforces the need for athlete monitoring in hockey to ensure that individual athletes are receiving appropriate training doses, given the lack of homogeneity across the sport. These data on the physical demands of hockey competition, as summarized in chapter 5 can also be used to evaluate athletes' performances and design in-season and off-season training programs (Gabbett, 2010). The absence of standardized thresholds for speed and acceleration/deceleration prevents comparison across the literature and between various hockey populations, limiting the implementation of reported data across populations. This limitation impacts the usefulness of monitoring data to create training programs and, on an individual athlete level, to monitor an athlete competing for multiple teams (club and international). Standardized thresholds should be established and utilized in athlete monitoring systems implemented based on the results of this thesis.

In summary, external training load is measured in hockey using GPS or GNSS units. These units were found to be valid and reliable measures of distance and speed, but they do have a negative bias, particularly during short, high-speed movements with change of direction.

Distance-based measures are the most frequently monitored external training load metrics and provide intuitive information on athlete's physical output. In competition settings, using relative measures of workrate can allow for comparisons across athletes playing for different minutes. Acceleration and deceleration data may provide valuable insights on muscular load and performance, particularly as athletes fatigue. Standardized thresholds need to be established for speed and acceleration zones to allow for comparisons across studies. The available data on external training load in hockey as analyzed and reviewed in chapter 4 can be used to set targets and design training programs to best prepare athletes for the demands of hockey competition.

9.3.2 Internal training Load

As part of this thesis, two new measures for internal training load, piTRIMP2 and iTRIMP2, were developed and, with its greater external validity, the novel piTRIMP2 was found to be the preferred metrics for use in hockey (Chapter 8). As outlined in section 7.1, there are several different algorithms for calculating TRIMP that have been developed over time. These algorithms differ not only in the equations themselves but also in the resources required for implementation via individualized athlete testing and discrete or continuous summation software. The current gold standard for TRIMP is iTRIMP in which heart rate is continuously weighted based on an individuals' blood lactate vs heart rate reserve (BLvHR) curve (Manzi *et al.*, 2013). As shown in chapter 9, work from this thesis has demonstrated that BLvHR can be calculated either via the laboratory-based protocol or a new pitch-based hockey-specific protocol. The pitch-based protocol designed and employed in this body of work has the benefit of accessibility and ecological validity, but further research is needed across a larger group of athletes to evaluate if this protocol produces more accurate iTRIMP scores predictive of fitness change. Additionally, this research found that iTRIMPs have a stronger dose-response relationship when the secondary heart-rate reserve term is removed from the equation, termed as iTRIMP2 (laboratory protocol) and piTRIMP2 (pitch protocol) (Equation 8.4; Table 8.3). With the strongest relationship with other training load measures and fitness change overtime, the newly developed piTRIMP2 and iTRIMP2 provide an advancement to the established iTRIMP.

Although piTRIMP2 and iTRIMP2 are the preferred metrics and should be utilized when resources allow, it is important to consider other variations of TRIMP for teams with limited resources. The biggest barrier to the use of piTRIMP2 and iTRIMP2 is the requirement for individualized athlete testing with blood lactate measurement to determine athletes' BLvHR response (Gonzalez-Fimbres *et al.*, 2019). Although piTRIMP2 removes some of these barriers by allowing testing to be performed in an athlete's usual training environment, it still requires blood lactate measurement. Therefore, when individual athlete testing is not possible, the recommendation is to use a 'team' TRIMP algorithm, which approximates BLvHR response based on group means in a similar population (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019). For hockey, these algorithms can be derived from Stagno's TRIMP and female TRIMP for male and female athletes, respectively (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019). These newly proposed adapted algorithms are as follows.

Equation 9.1: Stagno's TRIMP Continuous

$$y = 0.1225e^{3.9434 \times HRR}$$

Equation 9.2: Female TRIMP Continuous

$$y = 0.1102e^{4.3913 \times HRR}$$

To align with the results of thesis for iTRIMP2/piTRIMP2, both of these equations do not have the additional heart rate reserve term included in similar continuous team TRIMP algorithms such as that proposed by Gonzalez-Fimbres *et al.* (2019). In instances where continuous calculations cannot be performed to due to software limitations, the discrete versions of fTRIMP, sTRIMP, iTRIMP, or piTRIMP can be used (Stagno, Thatcher and Van Someren, 2007; Konerth, 2019).

Heart rate metrics monitor internal training load in terms of the demands placed on the cardiovascular system. Although intermittent, hockey is primarily aerobic in nature, with male international athletes averaging 85% of their maximum heart rate during competition (Lythe and Kilding, 2011; Buglione *et al.*, 2013; Lythe and Kilding, 2013). As a result, cardiovascular load is an appropriate metric for internal load measurement (Stagno, Thatcher and Van Someren, 2007). Nevertheless, average heart rate should be avoided in favor of TRIMPs whenever possible due to the intermittent nature of hockey. In cases where heart rate monitoring is not possible or a secondary measure is preferred, session rating of perceived exertion (sRPE) can also be used as a subjective measure of internal training load (Foster *et al.*, 2001).

In alignment with the literature (Haddad *et al.*, 2017), the results of this thesis have demonstrated sRPE to be a useful measure of internal training load in hockey. Specifically, sRPE and TRIMPs measured during hockey training and competition were found to strongly correlate over a 9-week period (Table 8.2). Perfect correlations would not be expected as sRPE and TRIMPs are distinct in what they measure (exertion versus cardiovascular load). Although subjective in nature, RPEs have been repeatedly shown to be a valid measure of internal load in team sport athletes, with RPEs indicating potential sources for concern when not trending the same as other internal and external load measures (Haddad *et al.*, 2017). Additionally, as heart-rate measures are not appropriate for resistance training activities (Banister, 1991), if athletes are performing any strength-based training outside of their hockey programming, different monitoring techniques such as differential RPEs could be considered (McLaren *et al.*, 2017; Vanrenterghem *et al.*, 2017). Vanrenterghem *et al*. (2017) have suggested that monitoring internal load should be separated into physiological (cardiovascular) and biomechanical load. Thus, using differential sRPEs for breathlessness and muscular exertion may provide more insight into the various aspects of physiological load (McLaren *et al.*, 2017; Vanrenterghem *et al.*, 2017).

Internal training load in hockey can be considered in terms of RPE and TRIMP. When possible, TRIMP should be used to monitor internal cardiovascular load, with the novel laboratory-based iTRIMP2 and pitch-based piTRIMP2 being the preferred metrics. With its greater external validity and hockey specificity, the results of this research would suggest that piTRIMP2 is the best method for summarizing heart rate data into internal load scores in hockey. When individual monitoring is not possible, the newly adapted continuous derivations of sTRIMP and fTRIMP, without the additional heart rate reserve term provide a valuable alternative (Stagno, Thatcher and Van Someren, 2007; Gonzalez-Fimbres *et al.*, 2019; Konerth, 2019). Using sRPEs provides subjective information regarding athlete's internal load, particularly when differential sRPEs are used to quantify muscular exertion or when sRPEs are used in combination with TRIMPs (Foster *et al.*, 2001; McLaren *et al.*, 2017; Vanrenterghem *et al.*, 2017). However, care should be taken when using sRPEs in isolation due to their subjective nature and the ability of athletes to manipulate responses.

9.3.3 Recovery monitoring

As reviewed in detail in chapter 3, recovery monitoring evaluates how athletes are responding to a given training stimulus outside of a strictly sports-specific setting (Saw, Main and Gastin, 2016). The goal of recovery monitoring is to ensure that athletes maintain a balance between stress and recovery and to detect cases of under-recovery to avoid maladaptive states of non-functional overreaching and overtraining (Lambert and Borresen, 2006; Kellmann, 2010). There is no gold-standard measure of athlete recovery, and recovery metrics can be broken down into subjective and objective measures, with subjective measures having been shown to be more sensitive to changes in athlete wellbeing status than objective measures (Saw, Main and Gastin, 2016). Recovery measures reviewed as part of this thesis include POMS, DALDA, RESTQ-S, ARSS, SRSS, CMJ height, autonomic nervous system function, heart rate variability and recovery, and blood-based measures, with strengths and weaknesses of each outlined in chapter 3 (Morgan *et al.*, 1987b; Chambers *et al.*, 1998; Kellmann *et al.*, 2001; Rushall, 1990; Coutts *et al.*, 2007; Andersson *et al.*, 2008; Johnston *et al.*, 2013; Meeusen *et al.*, 2013; Bellenger *et al.*, 2016a; Kallus and Kellmann, 2016; Cadegiani and Kater, 2017; Nässi *et al.*, 2017a; Starling *et al.*, 2019).

A four-week study undertaken as part of this thesis demonstrated that the RESTQ-S scales of sport stress and general stress had a significantly nontrivial positive dose-response relationship with athlete training load (Table 5.2). This suggests the potential validity of these scales as recovery measures, although more research will be needed to determine if they can be used to distinguish those athletes with a maladaptive training response. Similarly, Krueger *et al.* (2019) reported perceived recovery measured via SRSS scores to decrease over the course of a hockey tournament, and a systematic review found that ARSS and SRSS outperformed other questionnaires as subjective recovery measures (Jeffries *et al.*, 2020). Although effective in a research context, RESTQ-S is limited in practicality due to its 76-question length (Kellmann and Kallus, 2001). Therefore, the 8-question SRSS is likely more appropriate to use as part of a regular monitoring protocol for athletes (Nässi *et al.*, 2017a). When using subjective measures such as SRSS, honest reporting from athletes is key to successful monitoring (Meeusen *et al.*, 2013; Kölling *et al.*, 2015). Thus, care should be taken when implementing questionnaires to minimize the risk of bias and ensure that athletes are aware of who can and cannot view their responses and how those responses will be used (Meeusen *et al.*, 2013; Kölling *et al.*, 2015).

Objective recovery measures are limited in their ability to reflect athlete recovery status, but may be useful alongside subjective measures (Saw, Main and Gastin, 2016). As evaluated in chapter 5, the results of this research found that CMJ height, a measure of neuromuscular fatigue, was not sensitive to changes in training dose in hockey (Figure 5.2). As such this measure has little practical value or validity as a recovery measure in hockey. The review in chapter 3 found that the most commonly studied recovery metric in hockey was heart rate variability (HRV), assessed via an orthostatic test and measured in terms of changes in $LnRMSSD_{CV}$. Despite its correlation with TRIMP ($p < 0.01$), (Perrotta *et al.*, 2019a; González-Fimbres, Hernández-Cruz and Flatt, 2021), HRV has been shown not to differ between athletes entering maladaptive states of nonfunctional overreaching or overtraining and those responding positively to overload training (Meeusen *et al.*, 2013; Bellenger *et al.*, 2016a). Thus, when used in isolation, HRV has little practical benefit as a recovery monitoring measure. However, due to its sensitivity to athlete training dose and reflection of autonomic nervous system status (Perrotta *et al.*, 2019a; González-Fimbres, Hernández-Cruz and Flatt, 2021), HRV shows promise for use alongside subjective recovery measures such as RESTQ-S or SRSS. Notably, HRV and SRSS complement each other well, minimizing the weaknesses of the other measure, with the objective nature of HRV reducing the ability of athletes to manipulate responses and SRSS scores able to assist in determining when changes in HRV are likely to be indicative of positive or negative responses to increased load.

Overall, athlete recovery monitoring is key to promoting athlete wellbeing and performance by helping ensure that athletes balance stress and recovery and are positively responding to training. Subjective recovery measures have been shown to be more sensitive to changes in athlete recovery status than objective measures, but objective measures can still be considered alongside subjective measures (Saw, Main and Gastin, 2016). A short, valid, and practical questionnaire, SRSS is a good subjective measure of athlete recovery status in hockey (Krueger *et al.*, 2019; Jeffries *et al.*, 2020). When time allows, the longer RESTQ-S can also be used to evaluate subjective athlete recovery. Although limited when used in isolation due to its inability to distinguish between positive and maladaptive responses to overload, HRV may be a valuable complement to SRSS or RESTQ-S (Perrotta *et al.*, 2019a; González-Fimbres, Hernández-Cruz and Flatt, 2021). Thus, SRSS or RESTQ-S, used alongside HRV when resources allow, provides information on athlete recovery status and recovery-stress balance, and can be used to ensure that athletes maintain a positive response to training.

9.3.4 Fitness

Although not a load or recovery measure, another key component of an athlete monitoring model is an athlete's fitness level. Traditional definitions of physical fitness have been centered around the functional work capacity to complete a task (Pate, 1988). As described in detail in chapter 4, the work to be performed in hockey is dictated by the external and internal demands of competition with repeated high-intensity actions (running, sprinting, acceleration, deceleration, change of direction) performed with periods of low intensity work at an elevated heart rate. Thus, fitness can be thought of as an athletes' aerobic, anaerobic, and muscular conditioning that allows them to repeat the high-intensity actions required in the sport of hockey. Fitness can be measured in a variety of different ways with a large range of fitness assessments available for testing athletes. Tests vary in their intent with tests such as the 30-15 intermittent fitness test and submaximal blood lactate threshold test designed to assess athletes' maximal aerobic speed or lactate threshold (Buchheit, 2010; Stagno, Thatcher and Van Someren, 2007). On the other hand, assessments such as 300 yd shuttles target an athlete's anaerobic capacity and change of direction, while weightlifting and power assessments evaluate athletes' strength (Jones, 1991).

Incorporating change of direction and intermittent running, the 30-15 intermittent fitness test (30-15 IFT) is a valid and reliable measure of athlete fitness in intermittent ball sports such as hockey (Buchheit, 2010; Grgic, Lazinica and Pedisic, 2021; Stanković *et al.*, 2021). The test consists of 30 s of running followed immediately by 15 s of walking/resting over a 40 m length (Buchheit, 2010). The test begins at either 8 or 10 $km \cdot hr^{-1}$ and increases by 0.5 $km \cdot hr^{-1}$ each stage, with pace dictated by beeps on an audio file (Buchheit, 2010). Athletes run to volitional exhaustion or until they are no longer able to maintain the pace dictated by the beeps, with the velocity of the final stage completed taken as their score (Buchheit, 2010). The 30-15 IFT has been the subject of several reviews which have found it to be a valid and reliable measure of fitness for team sport athletes with good ecological validity (Buchheit, 2010; Grgic, Lazinica and Pedisic, 2021; Stanković *et al.*, 2021). The 30-15 IFT has been widely accepted and is frequently used both in the literature and in applied settings (Buchheit, 2010; Grgic, Lazinica and Pedisic, 2021; Stanković *et al.*, 2021). However, the 30-15 IFT is not without its limitations.

Being a maximal test, it can be largely influenced by athlete motivation levels, and it requires substantial effort to complete. Smaller changes in athlete fitness can be difficult to detect due to the substantial difference of 0.5 $km \cdot hr^{-1}$ between levels (Konerth, 2019). Additionally, with the running stages ending between lines in the testing setup, there is a notable learning effect with athletes familiarized with the testing protocol able to slow down prematurely and reduce energy expenditure.

Fitness testing is important in that it allows coaches to understand athletes' conditioning status and readiness to perform (Buchheit, 2008). This can be used to individualize and modify training demands to ensure that athletes are receiving the appropriate stimulus (Buchheit, 2008). Due to the maximal nature of most fitness assessments and the associated physical and mental demands, it is not appropriate to repeat these assessments too frequently and care should be taken when incorporating testing into training (Buchheit and Brown, 2020). Thus, in some instances, submaximal lactate threshold or heart rate recovery testing may be more appropriate when resources allow (Shushan *et al.*, 2022). Specifically, the pitch-based protocol newly developed as part of this thesis and outlined in detail in Chapter 7 could be used as an ecologically valid submaximal assessment for in-season monitoring. However, more research will be needed to validate this protocol as a measure of athlete fitness. Outside of testing, athletes' predicted fitness levels should be a continual consideration as part of an athlete monitoring system to ensure that individual athletes are improving/maintaining the fitness to perform when required.

9.3.5 Physical performance

A key aim of any athlete monitoring system is to maximize performance (Gabbett *et al.*, 2017). Unlike individual sports such as athletics and weightlifting, performance in intermittent-ball sports is much harder to quantify (Phillips *et al.*, 2010; Bangsbo, 2015; Costa *et al.*, 2022). The fittest athletes do not necessarily perform the best and the athletes who run the most do not necessarily make the best decisions. Therefore, a more nuanced approach must be taken when considering performance (Phillips *et al.*, 2010; Bangsbo, 2015; Costa *et al.*, 2022). Bangsbo (2015) argued that performance in the majority of sports can be broken down into physiological, psychological/social, tactical and technical components. This model of performance has been used in reviews of performance, talent development and selection frameworks (Phillips *et al.*, 2010; Costa *et al.*, 2022), and these four pillars were shown to be the key training areas in world and Olympic champions (Durand-Bush and Salmela, 2002). There is overlap across performance areas; for example, an athlete's recovery status and wellbeing may directly influence the psychological/social aspects of performance via their decision making. However, for the purpose of this analysis, only the physical performance component will be directly considered since it is most directly linked to athlete monitoring (Bangsbo, 2015).

Like overall performance, physical performance in hockey is not straightforward to measure. External load is the amount of physical work performed, so there is a natural relation to physical performance (Bangsbo, 2015; Ravé *et al.*, 2020). However, there is also the nuance that the best performing athletes are not necessarily the ones covering the most distance. For example, in hockey significant positional differences in external training load measures have been repeatedly reported (Table 4.4). Due to rolling substitutions, athletes also play for different amounts of time, both within and across positional groups (Polglaze *et al.*, 2015; Vescovi and Frayne, 2015; Vescovi, 2016; Casamichana *et al.*, 2018; McGuinness *et al.*, 2019; McMahon and Kennedy, 2019). Opposition matters, with athletes' loads increasing when playing teams at comparable competitive levels (Vinson, Gerrett and James, 2018). Furthermore, the amount of time a team spends in and out of possession and various tactical approaches influence athletes' external loads differently across positions (Konarski, Matuszynski and Strzelczyk, 2006; Cunniffe *et al.*, 2021). Even though physical performance is measured in terms of external training load, it is important to recognize the situational demands that may influence an athletes' external load. The best physically performing athletes in hockey may have the highest external load scores because in some situations doing so may not be tactically or positionally appropriate (Konarski, Matuszynski and Strzelczyk, 2006; Cunniffe *et al.*, 2021). Thus, physical performance was defined here as the *ability* to produce high physical outputs when required in competition. This does not mean that the athlete will necessarily always do more work but rather that they have the capacity to do so whenever the situation requires. Therefore, an aim of athlete monitoring is to better prepare athletes to increase their physical performance by having the ability to produce high amounts of physical work when needed in competition.

9.4 Practical applications

Athlete monitoring is critical in hockey due to the large variety in internal and external load in hockey athletes (Konerth, 2019; Chapter 5). For example, Figures 4.3 – 5.6 illustrate the wide range in total distance and workrate across international hockey athletes. Even within the same team of athletes completing the same training sessions and matches, this thesis demonstrated a weekly standard deviation in TRIMP of 356 AU, given a group mean value of 777 AU, and a standard deviation of 8.3 km given a mean total distance of 21.6 km (Table 5.1). With external and internal training load also shown to significantly differ across playing positions and between oppositions and tactical strategies, the demands of hockey are clearly varied (Chapter 5). It has been shown that without athlete monitoring, coaches are not able to accurately determine the work performed or recovery status of athletes (Bompa, 1999; Brink *et al.*, 2014; Kraft *et al.*, 2018). Thus, with athletes' load varying and coaches unable to predict these differences, athlete monitoring is critical in hockey.

The advantage of this new evidence-based model lies in its practical application. Figure 9.2 provides evidence-based recommendations for each monitoring category and their interactions, allowing practitioners to determine what data to collect based on their teams' individual situation and the resources available. Although beyond the scope of the present research, this model provides the framework for an athlete monitoring system to examine the relationship between these metrics and flag when misalignments occur. For example, a traffic light approach can be implemented to summarize athlete status (Robertson, Bartlett and Gastin, 2017; Thornton *et al.*, 2019). Thus, figure 9.2 can be used to select the evidence-based components for use in a hockey athlete monitoring system. The interactions between measures also provides a simple system to understand the relationship between variables. For example, three potential scenarios and their interpretations based on this model are outlined below.

1. A training session that consisted of small-sided games produced high TRIMP scores across the team but low total distance and high-speed meters. SRSS scores indicate poor recovery. The 30-15 IFT was performed the week prior, and the majority of the team completed the target level. Recommendation: Instead of pushing athletes harder in the subsequent training session to make up for the external load not achieved in the previous training, athletes should be rested as the high internal training load and poor recovery both indicate high training stress on the athletes, with good fitness scores suggesting poor athlete fitness was not the cause. The small-sided games likely meant that athletes did not cover as much distance due to the decreased playing area, while still maintaining a high-intensity, as indicated by internal

training load. This highlights the importance of measuring both internal and external training load, and coaches should adjust for the increased internal demands when implementing similar small-sided games in future training sessions.

2. After the injury of another athlete, Athlete A became a key player for the team and began playing more minutes than other athletes. Athlete A's weekly total distance and high-speed running distance is significantly higher than team averages and their own weekly load prior to the injury. However, Athlete A's iTRIMP and sRPEs are comparable to the team average (previously were below average). The team has not performed submaximal lactate threshold testing recently, but Athlete A was among the top group of performers at the start of the season. Athlete A's RESTQ-S scores show good sport recovery and low general stress and have actually improved as their role in the team has increased. Recommendation: Despite the increase in Athlete A's load, they are responding well to the increased training demands. Likely as a result of their high fitness and potentially due to outside life factors, they are able to manage the increased external training demands. There is no need to rest Athlete A and reduce their external training load so that it is in alignment with the group. However, their internal training load and recovery scores should be closely monitored to ensure that they are rested if any maladaptation begins to occur.
3. Athlete B performed poorly on the 30-15 IFT at the start of the season. It is now three weeks into the season, and Athlete B's iTRIMP scores are consistently increasing compared to the group. Their SRSS has indicated high stress and low recovery. The coach has suggested that Athlete B do additional running to work on their fitness because their workrate in competition is below the team's standards. Recommendation: The athlete should not perform additional running because they are already not responding well to the demands of the season. Likely a result of their poor fitness, Athlete B is not able to perform the same external training load as other athletes without significantly elevated physical demands. It is likely that the athlete is already overreaching and pushing them further would only lead to overtraining and further performance declines. The coach should be reassured that due to their elevated internal training load, Athlete B's fitness will already be improving through their regular training, and it is important that the athlete rest to allow for positive adaptation to occur.

With only a few components and a clear menu of monitoring options, the model developed through work in this thesis is easily accessible and provides a framework for practitioners to communicate decisions with coaches and/or athletes. It is designed to streamline data collection and the analysis of interactions by providing a range of validated metrics to allow for multifaceted monitoring while avoiding redundancy. This newly developed evidence-based model can be implemented to optimize athlete monitoring in hockey, improving athlete wellbeing and performance.

Chapter 10: Conclusion

10.1 Research questions

The aim of this thesis was to develop an evidence-based model for athlete monitoring hockey. With relevance to this aim, the conclusions to the research questions are addressed below.

1. *How does athlete recovery status relate to training load and performance?*

Athlete recovery status is determined in part by training load and is indicative of athletes' wellbeing and performance. As training load provides physiological stress on the athlete, to be a valid recovery measure, a metric should be sensitive to changes in training load. However, athlete recovery is also impacted by outside lifestyle and individual factors, so recovery measures should distinguish between athletes positively and negatively responding to training to detect under-recovery before performance declines occur. As part of this thesis, RESTQ-S and CMJ height were investigated as recovery monitoring measures, with RESTQ-S, particularly the subscales of sport stress and general stress, showing a significantly nontrivial dose-response relationship with training load. In contrast, CMJ height is limited in its validity as a training load measure in hockey, with no relation to training load. There is also evidence to suggest the shorter SRSS and measures of heart rate variability may be sensitive to changes in training load in hockey. Future research is needed in order to determine if RESTQ-S and SRSS can be used to distinguish athletes entering maladaptive training states.

2. *What are the physical and physiological demands of hockey competition across male and female, elite and sub-elite populations?*

The demands of hockey notably vary across populations, positions, and sexes. The results of the systematic review and meta-analysis undertaken during this research program demonstrated that elite male athletes covered significantly more total distance in hockey than elite female athletes. Workrate was also higher in elite male athletes than elite female athletes, with midfielders and forwards typically having higher workloads than defenders. Other measures of external load could not be compared across studies due to the difference in thresholds values or limited data; however, the results suggested that hockey athletes cover more distance at lower speeds (less than 15 $km \cdot hr^{-1}$), with distance in higher speed zones decreasing as speed increases. Meta-analyses were not performed on sub-elite populations due to the limited data and varied populations, but these groups typically had reduced loads in comparison to elite athletes. Mean heart rate in hockey was reported to average 85% maximum heart rate in male international athletes, with results ranging from 79-89% in female athletes. Future research should address the lack of clear definitions of internal and external training load variables, particularly the inconsistency of speed, acceleration, and heart rate zone thresholds, to allow for comparisons across studies.

3. *What is the accuracy and precision of external training load monitoring using current Global Navigation Satellite System (GNSS) models?*

The majority of external training load monitoring in research on hockey has used devices developed by Catapult Sports or a subsidiary. The results of the hockey-specific validity study undertaken found Catapult Sports' Vector S7 to have good interunit reliability for distance and speed with a coefficient of variation of 0.3%. There was an overall mean negative bias of 2.8% for distance and speed which could be corrected for with a multiplicative factor of 1.0286, but biases increased during short shuttles with change of direction. The interunit reliability of the Vector S7 was better than previous models. The overall mean bias was increased, but bias on the short shuttles was comparable or improved. Future technological advancements should focus on making GNSS devices more accurate over short sprints with change of direction, and future research should consider the validity and interunit reliability of other output measures such as acceleration and deceleration.

4. *What training impulse protocols and algorithms are best suited to calculate internal training load from heart rate data?*

As part of this thesis, two new algorithms/protocols for measuring TRIMP, iTRIMP2 and piTRIMP2 were developed and found to be the preferred TRIMP measures in hockey. The measure iTRIMP2 is a novel adaptation of Manzi's iTRIMP with the additional heart rate reserve term removed from the equation, as this term decreased correlation with other training load measures and the dose-response relationship with fitness change. Similarly, piTRIMP2 is based on the same algorithm but individuals' heart rate versus blood lactate curves were determined by a newly designed, more ecologically valid pitch-based fitness test. This new fitness test increases accessibility and was designed to mirror the demands of hockey, incorporating frequent change of direction and speed. These measures were evaluated over nine-weeks of a competitive hockey season, and the variation in iTRIMP2 and piTRIMP2 were shown to explain 88% and 84% of the variation in athlete fitness levels. When resources are limited, other TRIMP algorithms such as continuous adaptions of female TRIMP and Stagno's TRIMP can be implemented. Future research should investigate piTRIMP2 across a larger sample of athletes and consider the effectiveness of the testing protocol as a fitness assessment.

The demands of hockey vary significantly across individuals, due in part to minutes played, playing position, tactical styles, competition level and opposition. As coaches cannot accurately predict these varied demands without monitoring measures, athlete monitoring is of particular importance in hockey populations to ensure that all athletes are receiving appropriate training doses. To address this need, an evidence-based model for athlete monitoring in hockey was developed based on the results of thesis and is provided in Figure 9.2. Incorporating a range of validated markers for each monitoring component, this model provides a framework that can be used to set up a monitoring system for hockey athletes. It also illustrates the macrolevel relationships between the various components which are important to consider when analyzing monitoring results. The model developed as part of this thesis can be used to optimize athlete monitoring, increase performance, and improve wellbeing in hockey athletes.

10.2 Study limitations

The limitations of the individual studies have been discussed separately in the previous chapters, but the key limitations of the data collected will be addressed here. Firstly, as a result of focusing on one team of hockey athletes, the studies had relatively small sample sizes. Sample size was naturally limited by the number of athletes competing for a team (16), and this was further exacerbated by individuals moving between teams due to form and injuries. Repeated measures helped alleviate this concern with many data points collected for each athlete; however, the lack of independence of these data limited statistical power (Bland and Altman, 1995). Furthermore, the use of a convenience sample of a single hockey team decreased generalizability. Ideally, a cross-sectional analysis would have been performed across a population of hockey athletes, with a potential design incorporating a stratified random sample of athletes across all teams in a league. However, in this instance, the resources required prohibited this type of analysis.

The effect of menstrual cycle was not considered in the female athletes participating in the study. This was due to the already small sample size and prevalence of hormonal contraceptive use across the studies' participants. The direct impact of menstrual cycle on athletic performance is not clearly defined, with several review studies demonstrating mixed results (Jonge, 2003; Oosthuyse and Bosch, 2010; Tsampoukos *et al.*, 2010). There is also the consideration of perceived versus actual impact of various menstrual stages (Bossi *et al.*, 2013; Ross *et al.*, 2017; McNulty *et al.*, 2023). For example, it has been shown that both naturally menstruating and oral pill contraceptive user have reported increased perceived recovery time post training and decreased perceived exercise performance while bleeding (McNulty *et al.*, 2023), but lactate threshold has been found not to differ across stages of the menstrual cycle (Bossi *et al.*, 2013; Ross *et al.*, 2017). Current best practices would suggest that participants be homogenous with respect to menstrual status (naturally menstruating, hormonal contraceptive users) and that athletes' menstrual cycle stage be incorporated in the research design (Elliott-Sale *et al.*, 2021).

Finally, all studies were observational in nature, with no manipulation of athlete training loads or recovery. Although this was a conscious choice not to interfere with athletes' regular training, this reduced the conclusions that could be drawn. For example, athlete training load was never directly manipulated, and, as a result, results were correlational in nature.

10.3 Future directions

The evidence-based model for athlete monitoring in hockey developed through this research program sets the foundation for the implementation of athlete monitoring systems in hockey. Specifically, athlete monitoring systems should be developed based on the new model introduced here and an experimental approach should be utilized to determine individualized target training load and recovery thresholds in hockey athletes. Although no singular load target will be best for all athletes, athlete monitoring systems should be used to determine the most appropriate loads for individual athletes within a team. The impact of load and recovery should be considered both in terms of athlete fitness change over time but also athletes' tactical, technical, and decision-making performance in competition, quantified via video analysis, to give a more wholistic view of athlete performance. The aim of these future studies should be to implement the athlete monitoring model introduced here to improve performance and the training/recovery status of hockey athletes. Athlete monitoring systems can be adapted based on the performance level and resources available making athlete monitoring accessible across different levels of performance and improving the wellbeing of hockey athletes and quality of hockey overall.

Appendix A: Covid-19 Academic Impact Statement

<table>
<tr><th colspan="4">Durham University
Covid-19 Academic Impact Statement
Postgraduate Research Thesis (Masters by Research/Doctoral Programmes)</th></tr>
<tr><td>Student Name</td><td colspan="2">Student ID number</td><td>Department</td></tr>
<tr><td>Natalie Konerth</td><td colspan="2">000843624</td><td>Sport and Exercise Science</td></tr>
<tr><td colspan="3">Did Covid-19 prevent or impede you from completing part of your research project as originally intended?</td><td><u>Yes</u> / No</td></tr>
<tr><td colspan="4">If 'Yes', please state what Covid-19 prevented you from doing (maximum 200 words). For example, limitations to the data set or other primary sources due to travel restrictions, inability to run/replicate certain experiments due to restricted access, cutting short aspects of research due to additional caring responsibilities etc.</td></tr>
<tr><td colspan="4">Covid-19 significantly impacted this research program as hockey was not permitted during the national lockdowns. Therefore, data collection was modified, delayed, and shortened to accommodate the ever-evolving restrictions on team sport. Instead of a season-long study evaluating and prescribing training load in hockey competition, two shorter studies on recovery monitoring and training impulse were performed. The recovery monitoring study (Chapter 5) took place from October - November 2020 between lockdowns and, as such, was most significantly impacted. The study was limited in timeframe (4 weeks rather than 10 weeks), sample size (participant dropout due to self-isolation), sex of participants (only female team studied to avoid contact across teams), and lack of post-testing (due to sudden lockdown). Additionally, all research had to be performed outside, with no laboratory-based recovery measures or fitness testing permitted, and no video analysis of athlete performance could be performed. The training impulse study (Chapter 8) was also impacted with a smaller final sample size due to increased injury and athlete movement across teams following a year away from competitive sport. Also, lab-protocols at the time did not allow for gas-exchange analyses.</td></tr>
<tr><td colspan="2">Please state the dates over which the impact occurred</td><td colspan="2">From: March 2020 To: December 2021</td></tr>
<tr><td colspan="4">Please use the space below to provide a brief statement (up to 500 words) on any choices you have made and actions you have taken in response to anything you were prevented from doing as identified above. For example, reduction in the scope of the research, changes to the research design or revised research questions.</td></tr>
<tr><td colspan="4">The original aim of this research program was to determine the most effective method of using tracking devices, fitness testing and recovery measures to monitor training load and prescribe training dose in male and female hockey athletes. The research was intended to evaluate how to optimize athlete training load using athlete monitoring to modify training dose prescription. Additionally, the planned study was aimed at male and female hockey athletes, so the similarities and differences across the sexes in regard to training dose measurement and prescription were going to be evaluated. Specifically, a season-long study was proposed with an observational first half incorporating regular internal training load, external training load and recovery monitoring as well as bi-weekly submaximal fitness monitoring. The second part of the study was then planned to be experimental in approach with individualized training doses prescribed and updated based on fitness and recovery monitoring.</td></tr>
</table>

The goal of this study was to evaluate optimizing training dose prescription and its impact on athlete wellbeing, fitness, and performance.

As COVID-19 began during the first year of study and team sport was prohibited for much of the second year when data collection was planned to occur, significant changes were made to the aim and research questions to allow research to continue. Firstly, the choice was made to focus more in depth on the various components of athlete monitoring as these could be studied discretely in distinct studies, able to be adjusted based on evolving restrictions. For example, data collection for the recovery monitoring study was performed outside with no laboratory access in a time when there were still notable restrictions on team sport. As an extended and consistent period of data collection on the same athletes would have been required to incorporate both the observational and experiment aspects of training dose prescription evaluation, and it was unclear if and when this level of normalcy would be possible, the decision was made to forgo this aspect of the original research program. However, this choice allowed for more in depth analysis of the various aspects of recovery monitoring, particularly the development of the pitch-based protocol for individualized training load monitoring. The decision was also made to limit the research focus for the post-covid studies to incorporate only female athletes due to ease of access and concerns around potential contact and sharing of equipment across teams. As a result, comparison across sexes was not possible, but research was still performed on female athletes, a less frequently monitored population, particularly in reference to load and recovery.

Overall, the focus of the research was changed from training dose prescription to the development of an athlete monitoring model to allow research to continue despite the impact of covid-19. Individual studies were adjusted in accordance with the guidance and restrictions in place at the time.

I declare that the work submitted with this form was completed to the best of my ability in the light of the impact of Covid-19 as described above.

Candidate signature	
Principal Supervisor signature	
Date	20/12/23

Appendix B: Sample Meta-Analysis Calculations for Overall Total Distance

Female		**Sample**	**Files**	**m**	**SD**	**SE**	**w**	**wT**	**w(T-Tbar)**	**w^2**	**w***	**W*T***			
Kim et al	2016	32		5270	644	113.84419	7.7E-05	0.40662	4.67767	6E-09	6.2E-06	0.03255		Tbar	5023.7786
Vescovi	2016	44		4351	1282	193.26877	2.7E-05	0.11648	12.1177	7.2E-10	5.4E-06	0.02336		c	0.0001603
McGuinness et al	2018	27	154	4847	583	112.1984	7.9E-05	0.38503	2.48248	6.3E-09	6.2E-06	0.03001		t-hat^2	148942.67
McMahon Kennedy	2019	19	400	5029	995	228.26866	1.9E-05	0.09651	0.00052	3.7E-10	5E-06	0.02501		Tbar*	5029.3711
Morencos et al.	2019	16	112	5687	905	226.25	2E-05	0.1111	8.5929	3.8E-10	5E-06	0.02842		v*	36093.632
							0.00022	1.11575	27.8713	1.4E-08	2.8E-05	0.13934		SE(T*)	189.98324
															424.81544
Male		**Sample**	**Files**	**m**	**s**	**SE**	**w**	**wT**	**w(T-Tbar)**	**w^2**	**w***	**W*T***			
Lythe & Kidling	2011	18	90?	6798	2009	473.52584	4.5E-06	0.03032	4.3045	2E-11	2.2E-06	0.01483		Tbar	5815.5621
White & MacFarlane	2013	16	73	5819	687	171.75	3.4E-05	0.19727	0.0004	1.1E-09	3.8E-06	0.02206		c	0.0002063
Polglaze et al.	2015	24	105	6095	938	191.46845	2.7E-05	0.16626	2.12998	7.4E-10	3.7E-06	0.02249		t-hat^2	234319.88
White & MacFarlane	2015a	16	75	5868	665	166.25	3.6E-05	0.21231	0.09949	1.3E-09	3.8E-06	0.0224		Tbar*	6027.0443
Buglione et al	2017	12	36	7062	1015	293.00526	1.2E-05	0.08226	18.0963	1.4E-10	3.1E-06	0.02206		v*	31985.327
Ishan et al	2017	12	72	5232	479	138.27539	5.2E-05	0.27364	17.8108	2.7E-09	3.9E-06	0.02064		SE(T*)	178.84442
Sunderland Edwards	2017	20	234	6603	1089	243.5078	1.7E-05	0.11136	10.457	2.8E-10	3.4E-06	0.02249			536.53326
Polglaze et al.	2018	16	92	5523	632	158	4E-05	0.22124	3.42864	1.6E-09	3.9E-06	0.0213			
Krueger et al.	2019	18	48	5839	997	234.99515	1.8E-05	0.10574	0.00995	3.3E-10	3.5E-06	0.02017			
							0.00024	1.40038	56.3371	8.3E-09	3.1E-05	0.18843			

		n	Mean	SD	Lower	Upper
Kim et al	2016	32	5270	644	5038	5502
Vescovi	2016	44	4351	1282	3961	4741
McGuinness et al	2018	27	4847	583	4616	5078
McMahon Kennedy	2019	19	5029	995	4549	5509
Morencos et al.	2019	16	5687	905	5205	6169
Overall			5029		4657	5401
		n	Mean	SD	Lower	Upper
Lythe & Kidling	2011	18	6798	2009	5799	7797
White & MacFarlane	2013	16	5819	687	5453	6185
Polglaze et al.	2015	24	6095	938	5699	6491
White & MacFarlane	2015a	16	5868	665	5514	6222
Buglione et al	2017	12	7062	1015	6417	7707
Ishan et al	2017	12	5232	479	4928	5536
Sunderland Edwards	2017	20	6603	1089	6093	7113
Polglaze et al.	2019	16	5523	632	5186	5860
Krueger et al.	2019	18	5839	997	5343	6335
Overall			6027		5677	6377

Appendix C: COVID-19 limitations in recovery monitoring study

The coronavirus pandemic significantly impacted the timing, design, and execution of the recovery monitoring study. This section will outline the changes that were made and the impact of these changes. The researcher acknowledges the notable limitations of the recovery monitoring study and the data collected. However, given the continually changing restrictions due to COVID-19, the study could not be carried out as initially intended. The results of the shortened study are still presented in the following chapter as the data provide some preliminary evidence and set the foundation for future research in this area.

One of the largest modifications to the recovery monitoring study as a result of the coronavirus pandemic was the quantity of data collected. This limitation is threefold, in the sample size, study length, and frequency of data collection. Beginning with sample size, when the study was initially designed, the plan was to include thirty hockey athletes (15 male and 15 female) from the men's and women's first teams at Durham University. However, to minimize potential contact between groups and due to the logistical challenges of additional travel with the men's team (many of whose matches required overnight stays), this research was restricted to only the women's first team. Out of the seventeen female athletes initially recruited to participate in the study, ten athletes had to complete a period of self-isolation at some point during data collection. As athletes were required to have at least three complete weeks of training load data to be included in the final data analysis, the final sample size was reduce from seventeen to ten athletes. Initially the study was designed to last for a period of approximately twelve weeks during the first half of the 2020 hockey season (late September – December). However, as a result of the regional tiered restrictions and then the third national lockdown, only four weeks of data were collected. This time-period significantly limits the interpretation of recovery monitoring data, particularly as overreaching and overtraining take time to develop; however, as organized sport was not permitted during the third national lockdown, this limitation was unavoidable. Finally, the frequency of data collected during the four weeks was also negatively impacted by the coronavirus pandemic. Specifically, the athletes in the study would normally compete in two matches per week; however, only one of the two leagues was taking place, limiting the amount of match data collected.

In addition to the limitations in the quantity of data collected, changes were also made to the measurement variables. Specifically, as outlined in the following chapter, the data collected in the recovery monitoring study included subjective and objective recovery measures as well as measures of internal and external training load. However, what was missing was a measure of athlete performance. Simply examining the association between training load and recovery monitoring measures is only one piece in determining the validity of recovery metrics. To be a useful marker of athlete recovery, a measure must also be able to distinguish between those athletes positively and negatively responding to a given training load. Without performance data, this distinction cannot be made and all that can be examined is the sensitivity of the dose-response relationship between training load and recovery markers. Consequently, although the results of the recovery monitoring study performed provide preliminary information as to the sensitivity of recovery markers to changes in training load, they cannot be used to determine whether the recovery markers distinguish those entering a maladaptive training state. Prior to coronavirus, the researcher considered included several measures of athlete performance in the recovery monitoring study to overcome this limitation. Specifically, weekly submaximal fitness tests were planned to examine how athlete fitness levels were associated with changes in recovery markers. Unfortunately, these tests were not able to be included due to a decrease in the allotted training times as a result of social distancing and washing of equipment between training groups. The researcher had also considered monitoring athlete performance in competition via the tracking of unforced errors post-match using game-film, as a sport-specific decrease in performance is a primary indicator of non-functional overreaching and overtraining syndrome. However, as home matches were filmed from an indoor location and there was restricted access to the filming equipment, many of the matches were not filmed and this analysis could not be performed. When the study began, the plan was still to include pre-study and post-study 30-15 intermittent fitness tests, but the post-testing could not be completed due to the third national lockdown.

Despite these many limitations, data was collected on the dose-response relationship between RESTQ-S, CMJ height and training load measures. Future studies into recovery monitoring in hockey should include a larger sample size, monitored for a longer period, and should consider performance metrics to determine if these recovery measures can provide an early indication of athletes entering a maladaptive training state. Despite the increased research on recovery monitoring in hockey, no studies have evaluated recovery monitoring measures from this perspective. Therefore, more research is still needed to determine the most effective measures of monitoring athlete recovery in hockey athletes.

Appendix D: Prescreening Questionnaire

Prescreening Questionnaire

Adapted Physical Activities Readiness Questionnaire (PAR-Q)[1]

Name:____ Date:____________________

Please check **YES** or **NO** for each question below

1. Has your doctor ever said that you have heart trouble?

   YES **NO**

2. Do you frequently have pains in your heart and chest?

   YES **NO**

3. Do you often feel faint or have spells of severe dizziness?

   YES **NO**

4. Has your doctor said that your blood pressure is too high?

   YES **NO**

5. Has your doctor ever told you that you have a chronic bone or joint condition and should avoid high levels of activity?

   YES **NO**

6. Do you have any existing injuries for which you have not been cleared for regular activity by a physiotherapist or doctor?

   YES **NO**

7. Do you know of any good reason why you should not perform intense physical activity?

   YES **NO**

If you answered yes to any of the questions above, please explain below.

[1] Adapted from the Physical Fitness Readiness Questionnaire as outlined by Humphrey and Lakomy (2003).

Appendix E: Recovery Monitoring Consent Form

Project title: ***The relationship between athlete recovery, training load, and performance in elite hockey***

Researcher(s): Natalie Konerth
Department: Sport and Exercise Sciences
Contact details: natalie.m.konerth@durham.ac.uk

Supervisor name: Dr Karen Hind, Mr. Rob Cramb,
Supervisor contact details: karen.hind@durham.ac.uk, r.k.cramb@durham.ac.uk

This form is to confirm that you understand what is the purpose of the project, what is involved and that you are happy to take part. Please initial each box to indicate your agreement:

I confirm that I have read and understand the Information Sheet and the Privacy Notice for the above project.	
I have had sufficient time to consider the information and ask any questions I might have, and I am satisfied with the answers I have been given.	
I understand who will have access to personal data provided, how the data will be stored, and what will happen to the data at the end of the project.	
I agree to take part in the above project.	
I understand that my participation is entirely voluntary and that I am free to withdraw at any time without giving a reason.	
I understand the risks associated with participation in this project, including the risk of contracting COVID-19, and I agree to abide by all safety protocols put in place to mitigate this risk.	
I understand that anonymised (i.e. not identifiable) versions of my data may be archived and shared with others for legitimate research purposes.	
I give consent for my performance data (heart rate, positioning, jump height, and fitness test data) to be shared with members of Durham University Hockey Club's coaching staff.	

Participant's Signature______________________________ Date_____________

(NAME IN BLOCK LETTERS)____

Researcher's Signature_____________________________ Date_____________

(NAME IN BLOCK LETTERS)____

Appendix F: Validity and Reliability Consent Form

Project title: ***Validity and Interunit Reliability of Catapult Vector 10Hz Global Navigation Satellite System (GNSS) Units for Assessing Athlete Movement Patterns in Hockey***

Researcher(s): Natalie Konerth
Department: Sport and Exercise Sciences
Contact details: natalie.m.konerth@durham.ac.uk

Supervisor name: Dr Karen Hind, Mr. Rob Cramb,
Supervisor contact details: karen.hind@durham.ac.uk, r.k.cramb@durham.ac.uk

This form is to confirm that you understand what is the purpose of the project, what is involved and that you are happy to take part. Please initial each box to indicate your agreement:

I confirm that I have read and understand the Information Sheet dated [10/2/2020] and the Privacy Notice for the above project.	
I have had sufficient time to consider the information and ask any questions I might have, and I am satisfied with the answers I have been given.	
I understand who will have access to personal data provided, how the data will be stored, and what will happen to the data at the end of the project.	
I agree to take part in the above project.	
I understand that my participation is entirely voluntary and that I am free to withdraw at any time without giving a reason.	
I understand that anonymised (i.e. not identifiable) versions of my data may be archived and shared with others for legitimate research purposes.	

Participant's Signature______________________________ Date_____________

(NAME IN BLOCK LETTERS)____

Researcher's Signature_____________________________ Date_____________

(NAME IN BLOCK LETTERS)____

Appendix G: Training Impulse Consent Form

Project title: ***A modified training impulse (TRIMP) for elite female hockey***

Researcher(s): Natalie Konerth
Department: Sport and Exercise Sciences
Contact details: natalie.m.konerth@durham.ac.uk

Supervisor name: Dr Karen Hind, Mr. Rob Cramb,
Supervisor contact details: karen.hind@durham.ac.uk, r.k.cramb@durham.ac.uk

This form is to confirm that you understand what is the purpose of the project, what is involved and that you are happy to take part. Please initial each box to indicate your agreement:

I confirm that I have read and understand the Information Sheet and the Privacy Notice for the above project.	
I have had sufficient time to consider the information and ask any questions I might have, and I am satisfied with the answers I have been given.	
I understand who will have access to personal data provided, how the data will be stored, and what will happen to the data at the end of the project.	
I agree to take part in the above project.	
I understand that my participation is entirely voluntary and that I am free to withdraw at any time without giving a reason.	
I understand the risks associated with participation in this project, including the risk of contracting COVID-19, and I agree to abide by all safety protocols put in place to mitigate this risk.	
I understand that anonymised (i.e. not identifiable) versions of my data may be archived and shared with others for legitimate research purposes.	
I give consent for my performance data (heart rate, positioning, and fitness test data) to be shared with members of Durham University Hockey Club's coaching staff.	

Participant's Signature______________________________ Date_____________

(NAME IN BLOCK LETTERS)____

Researcher's Signature_____________________________ Date_____________

(NAME IN BLOCK LETTERS)____

References

Abbott, H. (2016) *Positional and match action profiles of elite women's field hockey players in relationship to the 2015 FIH rule changes.* Doctor of Physiology in Sport Physiology and Performance, East Tennessee State University, ProQuest Dissertations Publishing.

Abian, P., Del Coso, J., Salinero, J. J., Gallo-Salazar, C., Areces, F., Ruiz-Vicente, D., Lara, B., Soriano, L., Muñoz, V. and Abian-Vicen, J. (2015) 'The ingestion of a caffeinated energy drink improves jump performance and activity patterns in elite badminton players', *Journal of Sports Sciences,* 33(10), pp. 1042-1050.

Abt, G., Jobson, S., Morin, J.-B., Passfield, L., Sampaio, J., Sunderland, C. and Twist, C. (2022) 'Raising the bar in sports performance research', *Journal of Sports Sciences,* 40(2), pp. 125-129.

Akubat, I. and Abt, G. (2011) 'Intermittent exercise alters the heart rate–blood lactate relationship used for calculating the training impulse (TRIMP) in team sport players', *Journal of Science and Medicine in Sport,* 14(3), pp. 249-253.

Akubat, I., Patel, E., Barrett, S. and Abt, G. (2012) 'Methods of monitoring the training and match load and their relationship to changes in fitness in professional youth soccer players', *Journal of Sports Sciences,* 30(14), pp. 1473-1480.

Altmann, V., Hart, A., Vanlandewijck, Y., van Limbeek, J. and van Hooff, M. (2015) 'The impact of trunk impairment on performance of wheelchair activities with a focus on wheelchair court sports: A systematic review', *Sports Medicine - Open,* 1(22), pp. 1-14.

Andersson, H., Raastad, T., Nilsson, J., Paulsen, G., Garthe, I. and Kadi, F. (2008) 'Neuromuscular fatigue and recovery in elite female soccer: effects of active recovery', *Medicine &Amp; Science In Sports &Amp; Exercise,* 40(2), pp. 372-380.

Andrade, R., Wik, E. H., Rebelo-Marques, A., Blanch, P., Whiteley, R., Espregueira-Mendes, J. and Gabbett, T. J. (2020) 'Is the Acute: Chronic Workload Ratio (ACWR) associated with risk of time-loss injury in professional team sports? A systematic review of methodology, variables and injury risk in practical situations', *Sports Medicine,* 50(9), pp. 1613-1635.

Archbold, H. A. P., Rankin, A. T., Webb, M., Nicholas, R., Eames, N. W. A., Wilson, R. K., Henderson, L. A., Heyes, G. J., Davies, R. and Bleakley, C. M. (2018) 'Recurrent injury patterns in adolescent rugby', *Physical Therapy in Sport,* 33, pp. 12-17.

Astorino, T. A., Tam, P. A., Rietschel, J. C., Johnson, S. M. and Freedman, T. P. (2004) 'Changes in physical fitness parameters during a competitive field hockey season', *The Journal of Strength and Conditioning Research,* 18(4), pp. 850-854.

Atkinson, G. and Nevill, A. M. (2001) 'Selected issues in the design and analysis of sport performance research', *Journal of sports sciences,* 19(10), pp. 811-827.

Atkinson, G. and Reilly, T. (1996) 'Circadian variation in sports performance', *Sports medicine,* 21, pp. 292-312.

Aubry, A., Hausswirth, C., Julien, L., Coutts, A. and Buchheit, M. (2015) 'The development of functional overreaching is associated with a faster heart rate recovery in endurance athletes', *PLoS One,* 10(10), pp. e0139754.

Bakdash, J. Z. and Marusich, L. R. (2017) 'Repeated measures correlation', *Frontiers in Psychology,* 8, pp. 456.

Bakdash, J. Z. and Marusich, L. R. (2021) 'rmcorr: Repeated Measures Correlation'. Available at: https://CRAN.R-project.org/package=rmcorr.

Bangsbo, J. (1994) *Fitness training in football: a scientific approach.* Copenhagen: August Krogh Institute, University of Copenhagen.

Bangsbo, J. (2015) 'Performance in sports – With specific emphasis on the effect of intensified training', *Scandinavian Journal of Medicine & Science in Sports,* 25(S4), pp. 88-99.

Bangsbo, J., Iaia, F. M. and Krustrup, P. (2008) 'The Yo-Yo Intermittent Recovery Test', *Sports Medicine,* 38(1), pp. 37-51.

Banister, E. W. (1991) 'Modeling elite athlete performance', in MacDougall, J.D., Wenger, H.A. and Green, H.J. (eds.) *Physiological Testing of the High-Performance Athlete.* 2nd ed. Champaign, Illinois: Canadian Association of Sport Sciences, pp. 403-424.

Barboza, S., Joseph, C., Nauta, J., Mechelen, W. and Verhagen, E. (2018) 'Injuries in field hockey players: A systematic review', *Sports Medicine,* 48(4), pp. 849-866.

Bartlett, J. D. and Drust, B. (2021) 'A framework for effective knowledge translation and performance delivery of Sport Scientists in professional sport', *European journal of sport science,* 21(11), pp. 1579-1587.

Baumert, M., Brechtel, L., Lock, J., Hermsdorf, M., Wolff, R., Baier, V. and Voss, A. (2006) 'Heart rate variability, blood pressure variability, and baroreflex sensitivity in overtrained athletes', *Clinical Journal of Sports Medicine,* 16(5), pp. 412-417.

Beato, M., Bartolini, D., Ghia, G. and Zamparo, P. (2016) 'Accuracy of a 10 Hz GPS unit in measuring shuttle velocity performed at different speeds and distances (5–20m)', *Journal of Human Kinetics,* 54(1), pp. 15-22.

Beato, M., Devereux, G. and Stiff, A. (2018) 'Validity and reliability of global positioning system units (STATSports Viper) for measuring distance and peak speed in sports', *Journal of strength and conditioning research,* 32(10), pp. 2831-2837.

Bekraoui, N., Boussaidi, L., Cazorla, G. and Leger, L. (2020) 'Oxygen uptake, heart rate, and lactate responses for continuous forward running and stop-and-go running with and without directional changes', *Journal of Strength and Conditioning Research,* 34(3), pp. 699-707.

Bellenger, C., Thomson, R., Robertson, E., Davison, K., Nelson, M., Karavirta, L. and Buckley, J. (2017) 'The effect of functional overreaching on parameters of autonomic heart rate regulation', *European Journal of Applied Physiology,* 117(3), pp. 541-550.

Bellenger, C. R., Fuller, J. T., Thomson, R. L., Davison, K., Robertson, E. Y. and Buckley, J. D. (2016a) 'Monitoring athletic training status through autonomic heart rate regulation: A systematic review and meta-analysis', *Sports Medicine,* 46(10), pp. 1461-1486.

Bellenger, C. R., Thomson, R. L., Howe, P. R. C., Karavirta, L. and Buckley, J. D. (2016b) 'Monitoring athletic training status using the maximal rate of heart rate increase', *Journal of Science and Medicine in Sport,* 19(7), pp. 590-595.

Berglund, B. and Safstrom, H. (1994) 'Psychological monitoring and modulation of training load of world-class canoeists', *Medicine & Science in Sports & Exercise,* 26(8), pp. 1036-1040.

Bishop, D. (2008) 'An applied research model for the sport sciences', *Sports Medicine,* 38(3), pp. 253-263.

Blanch, P. and Gabbett, T. (2015) 'Has the athlete trained enough to return to play safely? The acute:chronic workload ratio permits clinicians to quantify a player's risk of subsequent injury', *British journal of sports medicine,* 50(8), pp. 471-475.

Bland, J. M. and Altman, D. G. (1994) 'Statistics Notes: Correlation, regression, and repeated data', *British Medical Journal,* 308(6933), pp. 896.

Bland, J. M. and Altman, D. G. (1995) 'Calculating correlation coefficients with repeated observations: Part 1-- Correlation within subjects', *British Medical Journal (Clinical research ed.),* 310(6977), pp. 446.

Bompa, T. O. (1999) *Periodization: Theory and methodology of training.* 4th ed. edn. Champaign, IL: Human Kinetics.

Boran, A. (2012) *Comparison of distance travelled, speed and heart rate among three outfield positions in amateur female hockey players.* Masters of Science, University of Chester.

Borg, E. and Borg, G. (2002) 'A comparison of AME and CR100 for scaling perceived exertion', *Acta Psychologica,* 109(2), pp. 157-175.

Bosquet, L., Montpetit, J., Arvisais, D. and Mujika, I. (2007) 'Effects of tapering on performance: A meta-analysis', *Medicine & Science in Sports & Exercise,* 39(8), pp. 1358-1365.

Bosquet, L., Papelier, Y., Léger, L. and Legros, P. (2003) 'Night heart rate variability during overtraining in male endurance athletes', *The Journal of Sports Medicine and Physical Fitness,* 43(4), pp. 506.

Bossi, J., Kostelis, K., Walsh, S. and Sawyer, J. (2013) 'Effects of menstrual cycle on exercise in collegiate female athletes', *Research Quarterly for Exercise and Sport,* 84(S1), pp. AXXII.

Bourdon, P. C. (2017) 'Monitoring athlete training loads: Consensus statement', *International Journal of Sports Physiology & Performance,* 12, pp. S2-161.

Brink, M. S., Frencken, W. G., Jordet, G. and Lemmink, K. A. (2014) 'Coaches' and players' perceptions of training dose: Not a perfect match', *International Journal of Sports Physiology and Performance,* 9(3), pp. 497-502.

Bruce, L. M. and Moule, S. J. (2017) 'Validity of the 30-15 intermittent fitness test in subelite female athletes', *Journal of Strength and Conditioning Research,* 31(11), pp. 3077-3082.

Brun, J. F. (2003) 'Le surentraînement : à la recherche d'un outil d'évaluation standardisé utilisable en routine', *Science & sports,* 18(6), pp. 282-286.

Bryman, A. (1984) 'The debate about quantitative and qualitative research: A question of method or epistemology?', *The British Journal of Sociology,* 35(1), pp. 75-92.

Buchheit, M. (2008) 'The 30-15 intermittent fitness test: Accuracy for individualizing interval training of young intermittent sport players', *Journal of Strength and Conditioning Research,* 22(2), pp. 365-374.

Buchheit, M. (2010) 'The 30–15 intermittent fitness test: 10 year review', *Myorobie Journal,* 1(9), pp. 278-286.

Buchheit, M. (2014) 'Monitoring training status with HR measures: do all roads lead to Rome?', *Frontiers in Physiology,* 5, pp. 73.

Buchheit, M., Al Haddad, H., Millet, G. P., Lepretre, P. M., Newton, M. and Ahmaidi, S. (2009) 'Cardiorespiratory and cardiac autonomic responses to 30-15 intermittent fitness test in team sport players', *Journal of Strength and Conditioning Research,* 23(1), pp. 93-100.

Buchheit, M., Bishop D Fau - Haydar, B., Haydar B Fau - Nakamura, F. Y., Nakamura Fy Fau - Ahmaidi, S. and Ahmaidi, S. (2010) 'Physiological responses to shuttle repeated-sprint running', *International Journal of Sports Medicine,* 31(6), pp. 402-409.

Buchheit, M. and Brown, M. (2020) 'Pre-season fitness testing in elite soccer: Integrating the 30–15 Intermittent Fitness Test into the weekly microcycle', *Sport Performance and Science Reports, 1,* 111.

Buchheit, M., Haydar, B. and Ahmaidi, S. (2012) 'Repeated sprints with directional changes: Do angles matter?', *Journal of Sports Sciences,* 30(6), pp. 555-562.

Buchheit, M. and Simpson, B. M. (2017) 'Player-tracking technology: Half-full or half-empty glass?', *International Journal of Sports Physiology & Performance,* 12(Suppl 2), pp. S235-S241.

Buckworth, J. (2002) *Exercise psychology.* Leeds: Human Kinetics.

Buglione, A., Ruscello, B., Milia, R., Migliaccio, G. M., Granatelli, G. and D'Ottavio, S. (2013) 'Physical and physiological demands of elite and sub-elite field hockey players', *International Journal of Performance Analysis in Sport,* 13(3), pp. 872-884.

Burt, D., Hayman, O., Forsyth, J., Doma, K. and Twist, C. (2020) 'Monitoring indices of exercise-induced muscle damage and recovery in male field hockey: Is it time to retire creatine kinase?', *Science & Sports,* 35(6), pp. 402-404.

Busso, T., Denis, C., Bonnefoy, R., Geyssant, A. and Lacour, J. R. (1997) 'Modeling of adaptations to physical training by using a recursive least squares algorithm', *Journal of Applied Physiology,* 82(5), pp. 1685-1693.

Cadegiani, F. and Kater, C. (2017) 'Hormonal aspects of overtraining syndrome: A systematic review', *BMC Sports Science, Medicine & Rehabilitation,* 9(1), pp. 1-15.

Casamichana, D., Castellano, J. and Castagna, C. (2012) 'Comparing the physical demands of friendly matches and small-sided games in semiprofessional soccer players', *Journal of Strength and Conditioning Research,* 26(3), pp. 837-843.

Casamichana, D., Morencos, E., Romero-Moraleda, B. and Gabbett, T. J. (2018) 'The use of generic and individual speed thresholds for assessing the competitive demands of field hockey', *Journal of Sports Science and Medicine,* 17(3), pp. 366-371.

Castellano, J., Casamichana, D., Calleja-González, J. and Ostojic, S. (2011) 'Reliability and accuracy of 10 GPS devices for short- distance exercise', *Journal of Sports Science and Medicine,* 10(1), pp. 233-234.

Catapult Sports (2019) 'Vector Data Integrity', pp. 9-10, *Catapultsports.com*. Available at: https://support.catapultsports.com/hc/en-us/article_attachments/360002703476/Vector_Data_Validation.pdf (Accessed January 26, 2020).

Catapult Sports (2020) *Vector*. Available at: https://www.catapultsports.com/products/vector (Accessed: January 26, 2020).

Chalencon, S., Busso, T., Lacour, J.-R., Garet, M., Pichot, V., Connes, P., Gabel, C. P., Roche, F. and Barthelemy, J. C. (2012) 'A model for the training effects in swimming demonstrates a strong relationship between parasympathetic activity, performance and index of fatigue', *PLoS ONE,* 7(12), pp. e52636.

Chambers, C., Noakes, T. D., Lambert, E. V. and Lambert, M. I. (1998) 'Time course of recovery of vertical jump height and heart rate versus running speed after a 90- km foot race', *Journal of Sports Sciences,* 16(7), pp. 645-651.

Chesher, S. M., Netto, K. J., Appleby, B. B., Jacques, A. and Wild, C. Y. (2019) 'Deceleration characteristics of elite Australian male field hockey players during an Olympic tournament', *Journal of Science & Medicine in Sport,* 22(5), pp. 611-615.

Clark, M. E., McEwan, K. and Christie, C. J. (2019) 'The effectiveness of constraints-led training on skill development in interceptive sports: A systematic review', *International Journal of Sports Science & Coaching,* 14(2), pp. 229-240.

Claudino, J., Gabbett, T., Simim, M., Fowler, P., Melo, M., Bottino, A., Loturco, I., Amadio, A., Serrão, J. and Nassis, G. (2019) 'Which parameters to use for sleep quality monitoring in team sport athletes? A systematic review and meta-analysis', *BMJ Open Sport & Exercise Medicine,* 5(1).

Cormack, S. J., Newton, R. U., McGuigan, M. R. and Doyle, T. L. A. (2008) 'Reliability of measures obtained during single and repeated countermovement jumps', *International journal of sports physiology and performance,* 3(2), pp. 131-144.

Cormier, P., Freitas, T. T., Rubio-Arias, J. and Alcaraz, P. E. (2020) 'Complex and contrast training: Does strength and power training sequence affect performance-based adaptations in team sports? A Systematic review and meta-analysis', *Journal of Strength & Conditioning Research,* 34(5), pp. 1461-1479.

Costa, Y. P. d., Freitas-Júnior, C., Lima-Júnior, D. d., Soares-Silva, E. L., Batista, G. R., Hayes, L. and Fortes, L. d. S. (2022) 'Mental fatigue and ball sports: A narrative review focused on physical, technical, and tactical performance', *Motriz: Revista de Educação Física,* 28.

Coutts, A., Reaburn, P., Piva, T. J. and Murphy, A. (2007) 'Changes in selected biochemical, muscular strength, power, and endurance measures during deliberate overreaching and tapering in rugby league players', *International Journal of Sports Medicine,* 28(2), pp. 116-124.

Coutts, A. J., Crowcroft, S. and Kempton, T. (2017) 'Developing athlete monitoring systems: Theoretical basis and practical applications', in *Sport, Recovery, and Performance*. pp. 17-31 [Online]. Version.

Coutts, A. J., Slattery, K. M. and Wallace, L. K. (2007) 'Practical tests for monitoring performance, fatigue and recovery in triathletes', *Journal of Science and Medicine in Sport,* 10(6), pp. 372-381.

Coutts, A. J., Wallace, L. K. and Slattery, K. M. (2007) 'Monitoring changes in performance, physiology, biochemistry, and psychology during overreaching and recovery in triathletes', *International Journal of Sports Medicine,* 28(2), pp. 125-134.

Covic, N., Jeleskovic, E., Alic, H., Rado, I., Kafedzic, E., Sporis, G., McMaster, D. T. and Milanovic, Z. (2016) 'Reliability, validity and usefulness of 30-15 intermittent fitness test in female soccer players.', *Frontiers in Physiology,* 7, pp. 510.

Crang, Z. L., Duthie, G., Cole, M. H., Weakley, J., Hewitt, A. and Johnston, R. D. (2021) 'The validity and reliability of wearable microtechnology for intermittent team sports: A systematic review', *Sports Medicine,* 51(3), pp. 549-565.

Crewther, B. T., Hamilton, D., Casto, K., Kilduff, L. P. and Cook, C. J. (2015) 'Effects of oral contraceptive use on the salivary testosterone and cortisol responses to training sessions and competitions in elite women athletes', *Physiology & Behavior,* 147, pp. 84-90.

Crewther, B. T., Hamilton, D., Kilduff, L. P., Drawer, S. and Cook, C. J. (2018) 'The effect of oral contraceptive use on salivary testosterone concentrations and athlete performance during international field hockey matches', *Journal of Science and Medicine in Sport,* 21(5), pp. 453-456.

Cronin, J. and Hansen, K. T. (2006) 'Resisted sprint training for the acceleration phase of sprinting', *Strength and Conditioning Journal,* 28(4), pp. 42-51.

Cullen, B. D., Roantree, M. T., McCarren, A. L., Kelly, D. T., Connor, P. L., Hughes, S. M., Daly, P. G. and Moyna, N. M. (2017) 'Physiological profile and activity pattern of minor Gaelic football players', *Journal of Strength and Conditioning Research,* 31(7), pp. 1811-1820.

Cummins, C., Orr, R., O'Connor, H. and West, C. (2013) 'Global positioning systems (GPS) and microtechnology sensors in team sports: A systematic review', *Sports Medicine,* 43(10), pp. 1025-1042.

Cunniffe, E., Connor, M., Beato, M., Grainger, A., McConnell, W., McCarthy Persson, U., Delahunt, E., Boreham, C. and Blake, C. (2021) 'The influence of possession status on the physical output of male international hockey players', *International Journal of Sports Science & Coaching,* 17(2), pp. 412-422.

Daanen, H. A. M., Lamberts, R. P., Kallen, V. L., Jin, A. and Van Meeteren, N. L. U. (2012) 'A systematic review on heart-rate recovery to monitor changes in training status in athletes', *International journal of sports physiology and performance,* 7(3), pp. 251.

David, M. and Sutton, C. D. (2011) *Social Research : An Introduction.* 2nd ed. edn. Los Angeles: Los Angeles : SAGE.

Davis, H., Orzeck, T. and Keelan, P. (2007) 'Psychometric item evaluations of the Recovery-Stress Questionnaire for athletes', *Psychology of Sport & Exercise,* 8(6), pp. 917-938.

Delextrat, A., Trochym, E. and Calleja-Gonzalez, J. (2012) 'Effect of a typical in-season week on strength jump and sprint performances national-level female basketball players', *The Journal of Sports Medicine and Physical Fitness,* 52(2), pp. 128-136.

Dellal, A., Keller, D., Carling, C., Chaouachi, A., Wong, D. P. and Chamari, K. (2010) 'Physiologic effects of directional changes in intermittent exercise in soccer players', *Journal of Strength and Conditioning Research,* 24(12), pp. 3219-3226.

Dobbs, C. W., Gill, N. D., Smart, D. J. and McGuigan, M. R. (2015) 'Relationship between vertical and horizontal jump variables and muscular performance in athletes', *Journal of Strength and Conditioning Research,* 29(3), pp. 661.

Doherty, C., Delahunt, E., Caulfield, B., Hertel, J., Ryan, J. and Bleakley, C. (2014) 'The incidence and prevalence of ankle sprain injury: A systematic review and meta- analysis of prospective epidemiological studies', *Sports Medicine,* 44(1), pp. 123-140.

Drew, M. and Finch, C. (2016) 'The Relationship Between Training Load and Injury, Illness and Soreness: A Systematic and Literature Review', *Sports Medicine,* 46(6), pp. 861-883.

Drust, B. and Green, M. (2013) 'Science and football: evaluating the influence of science on performance', *Journal of sports sciences,* 31(13), pp. 1377-1382.

Duffield, R., Bosquet, L., Erlacher, D., Robazza, C., Halson, S. L., Brink, M., Heidari, J., Venter, R., Meeusen, R., Kellmann, M., Bertollo, M., Coutts, A. J., Hecksteden, A., Mujika, I., Beckmann, J., Kallus, K. W. and Skorski, S. (2018) 'Recovery and Performance in Sport: Consensus Statement', *International Journal of Sports Physiology & Performance,* 13(2), pp. 240-245.

Dupuy, O., Bherer, L., Audiffren, M. and Bosquet, L. (2013) 'Night and postexercise cardiac autonomic control in functional overreaching', *Applied Physiology, Nutrition, and Metabolism,* 38(2), pp. 200-208.

Durand-Bush, N. and Salmela, J. H. (2002) 'The development and maintenance of expert athletic performance: perceptions of world and olympic champions', *Journal of Applied Sport Psychology,* 14(3), pp. 154-171.

Dwyer, D. B. and Gabbett, T. J. (2012) 'Global positioning system data analysis: velocity ranges and a new definition of sprinting for field sport athletes', *Journal of Strength and Conditioning Research,* 26(3), pp. 818.

Eckard, T., Padua, D., Hearn, D., Pexa, B. and Frank, B. (2018) 'The relationship between training load and injury in athletes: A systematic review', *Sports Medicine,* 48(8), pp. 1929-1961.

Edwards, S. (1993) *The Heart Rate Monitor Book.* 2nd edn.: Polar Electro Inc.

Eichner, E. R. (1995) 'Overtraining: consequences and prevention', *Journal of Sports Sciences,* 13(Sup 1), pp. S41-S48.

Elliott-Sale, K. J., Minahan, C. L., de Jonge, X. A. K. J., Ackerman, K. E., Sipilä, S., Constantini, N. W., Lebrun, C. M. and Hackney, A. C. (2021) 'Methodological considerations for studies in sport and exercise science with women as participants: A working guide for standards of practice for research on women', *Sports Medicine,* 51(5), pp. 843-861.

Enders, C. K. and Tofighi, D. (2007) 'Centering predictor variables in cross-sectional multilevel models: a new look at an old issue', *Psychological Methods,* 12, pp. 121-138.

England Hockey (2023) *About England Hockey*. Available at: https://www.englandhockey.co.uk/governance/about-england-hockey (Accessed: August 26, 2023.

Filaire, E., Bernain, X., Sagnol, M. and Lac, G. (2001) 'Preliminary results on mood state, salivary testosterone:cortisol ratio and team performance in a professional soccer team', *European Journal of Applied Physiology,* 86(2), pp. 179-84.

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P. and Dodge, C. (2001) 'A new approach to monitoring exercise training', *Journal of Strength and Conditioning Research,* 15(1), pp. 109-116.

Fowles, J. R. (2006) 'Technical issues in quantifying low- frequency fatigue in athletes', *International Journal of Sports Physiology and Performance,* 1(2), pp. 169-171.

Fox, J., Stanton, R., Sargent, C., Wintour, S.-A. and Scanlan, A. (2018) 'The association between training load and performance in team sports: A systematic review', *Sports Medicine,* 48(12), pp. 2743-2774.

Fox, J. L., Scanlan, A. T. and Stanton, R. (2017) 'A review of player monitoring approaches in basketball: Current trends and future directions', *The Journal of Strength & Conditioning Research,* 31(7).

Freitas, V., Nakamura, F., Miloski, B., Samulski, D. and Bara-Filho, M. (2014) 'Sensitivity of physiological and psychological markers to training load intensification in volleyball players', *Journal of Sports Science & Medicine,* 13(3), pp. 571-579.

Gabbett, T. J. (2010) 'GPS analysis of elite women's field hockey training and competition', *Journal of Strength and Conditioning Research,* 24(5), pp. 1321-1324.

Gabbett, T. J. (2016) 'The training—injury prevention paradox: should athletes be training smarter *and* harder?', *British Journal of Sports Medicine,* 50(5), pp. 273-280.

Gabbett, T. J. and Domrow, N. (2007) 'Relationships between training load, injury, and fitness in sub-elite collision sport athletes', *Journal of Sports Sciences,* 25(13), pp. 1507-1519.

Gabbett, T. J., Nassis, G. P., Oetter, E., Pretorius, J., Johnston, N., Medina, D., Rodas, G., Myslinski, T., Howells, D., Beard, A. and Ryan, A. (2017) 'The athlete monitoring cycle: a practical guide to interpreting and applying training monitoring data', *British Journal of Sports Medicine,* 51(20), pp. 1451-1452.

Gastin, P., Hunkin, S., Fahrner, B. and Robertson, S. (2019) 'Deceleration, acceleration, and impacts are Strong contributors to muscle damage in professional australian football', *Journal of Strength and Conditioning Research,* 33(12), pp. 3374-3383.

Glatthorn, J. F., Gouge, S., Nussbaumer, S., Stauffacher, S., Impellizzeri, F. M. and Maffiuletti, N. A. (2011) 'Validity and reliability of Optojump photoelectric cells for estimating vertical jump height', *Journal of Strength and Conditioning Research,* 25(2), pp. 556-560.

Gleeson, M. (2002) 'Biochemical and immunological markers of over-training', *Journal of Sports Science and Medicine,* 1(2), pp. 31-41.

Gløersen, Ø., Kocbach, J. and Gilgien, M. (2018) 'Tracking performance in endurance racing sports: evaluation of the accuracy offered by three commercial GNSS receivers aimed at the sports market', *Frontiers in Physiology,* 9(1425), pp. 1425.

González-Fimbres, R. A., Hernández-Cruz, G. and Flatt, A. A. (2021) 'Ultrashort versus criterion heart rate variability among international-level girls' field hockey players', *International Journal of Sports Physiology and Performance,* 16(7), pp. 985-992.

Gonzalez-Fimbres, R. A., Ramirez-Siqueiros, M. G., Reynoso-Sanchez, L. F., Quezada-Chacon, J. T., Miranda-Mendoza, J. and Hernandez-Cruz, G. (2019) 'A new approach to quantify internal and external training load for intermittent sports', *Biotecnia,* 21(3), pp. 26-34.

Gould, D., Greenleaf, C., Chung, Y. and Guinan, D. (2002) 'A survey of U.S. Atlanta and Nagano olympians: variables perceived to influence performance', *Research Quarterly for Exercise and Sport,* 73(2), pp. 175-186.

Gratton, C. (2010) *Research Methods for Sports Studies.* 2nd ed. edn. London: Routledge.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N. and Altman, D. G. (2016) 'Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations', *European journal of epidemiology,* 31(4), pp. 337-350.

Grgic, J., Lazinica, B. and Pedisic, Z. (2021) 'Test–retest reliability of the 30–15 Intermittent Fitness Test: A systematic review', *Journal of Sport and Health Science,* 10(4), pp. 413-418.

Gutmann, M. C., Pollock, M. L., Foster, C. and Schmidt, D. (1984) 'Training stress in olympic speed skaters: A psychological perspective', *The Physician and Sportsmedicine,* 12(12), pp. 45-57.

Haddad, M., Stylianides, G., Djaoui, L., Dellal, A. and Chamari, K. (2017) 'Session-RPE method for training load monitoring: validity, ecological usefulness, and influencing factors', *Frontiers in Neuroscience,* 11, pp. 612.

Hagstrom, A. D. and Shorter, K. A. (2018) 'Creatine kinase, neuromuscular fatigue, and the contact codes of football: A systematic review and meta-analysis of pre- and post-match differences', *European Journal of Sport Science,* 18(9), pp. 1234-1244.

Halson, S. (2014) 'Monitoring training load to understand fatigue in athletes', *Sports Medicine,* 44(Supplement 2), pp. 139-147.

Halson, S. L., Bridge, M. W., Meeusen, R., Busschaert, B., Gleeson, M., Jones, D. A. and Jeukendrup, A. E. (2002) 'Time course of performance changes and fatigue markers during intensified training in trained cyclists', *Journal of Applied Physiology,* 93(3), pp. 947-56.

Hamilton, D. 2019. Field Hockey. *In:* Laursen, P. and Buchheit, M. (eds.) *Science and application of high-intensity interval training : solutions to the programming puzzle.* Champaign, IL : Human Kinetics.

Hamlin, M. J., Wilkes, D., Elliot, C. A., Lizamore, C. A. and Kathiravel, Y. (2019) 'Monitoring training loads and perceived stress in young elite university athletes', *Frontiers in Physiology,* 10, pp. 34.

Hammes, D., Skorski, S., Schwindling, S., Ferrauti, A., Pfeiffer, M., Kellmann, M. and Meyer, T. (2016) 'Can the Lamberts and Lambert Submaximal Cycle Test indicate fatigue and recovery in trained cyclists?', 11(3), pp. 328-336.
Harper, D. J., Carling, C. and Kiely, J. (2019) 'High-intensity acceleration and deceleration demands in elite team sports competitive match play: A systematic review and meta-analysis of observational studies', *Sports Medicine,* 49(12), pp. 1923-1947.
Haydt, R., Pheasant, S. and Lawrence, K. (2012) 'The incidence of low back pain in ncaa division III female field hockey players', *International journal of sports physical therapy,* 7(3), pp. 296.
Hedelin, R., Goran, K., Wiklund, U., Bjerle, P. and Henriksson-Larsen, K. (2000) 'Short-term overtraining: effects on performance, circulatory responses, and heart rate variability', *Medicine & Science in Sports & Exercise,* 32(8), pp. 1480-1484.
Hedges, L. V. and Vevea, J. L. (1998) 'Fixed- and Random-Effects Models in Meta-Analysis', *Psychological Methods,* 3(4), pp. 486-504.
Heidari, J., Beckmann, J., Bertollo, M., Brink, M., Kallus, K. W., Robazza, C. and Kellmann, M. (2019) 'Multidimensional monitoring of recovery status and implications for performance', *International Journal of Sports Physiology and Performance,* 14(1), pp. 2-8.
Heishman, A. D., Daub, B. D., Miller, R. M., Freitas, E. D. S., Frantz, B. A. and Bemben, M. G. (2020) 'Countermovement jump reliability performed with and without an arm swing in NCAA Division 1 intercollegiate basketball players', *Journal of Strength and Conditioning Research,* 34(2), pp. 546-558.
Henderson, M. (2020) *How to tidy Catapult 10Hz export data*. mitchhenderson.org. Available at: https://www.mitchhenderson.org/2020/04/how-to-tidy-catapult-10hz-export-data/ 2022).
Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. and Welch, V. e. (2019) *Cochrane handbook for systematic reviews of interventions. Cochrane handbook for systematic reviews of interventions*: Cochrane. Available at: www.training.cochrane.org/handbook.
Hopkins, W. (2000) 'Measures of Reliability in Sports Medicine and Science', *Sports Medicine,* 30(1), pp. 1-15.
Hoppe, M. W., Baumgart, C., Slomka, M., Polglaze, T. and Freiwald, J. (2017) 'Variability of metabolic power data in elite soccer players during pre-season matches', *Journal of Human Kinetics,* 58(1), pp. 233-245.
Ihsan, M., Tan, F., Sahrom, S., Choo, H. C., Chia, M. and Aziz, A. R. (2017) 'Pre-game perceived wellness highly associates with match running performances during an international field hockey tournament', *European Journal of Sport Science,* 17(5), pp. 593-602.
Impellizzeri, F. M., Marcora, S. M. and Coutts, A. J. (2022) 'Internal and External Training Load: 15 Years On', *International Journal of Sports Physiology and Performance,* 14(2), pp. 270-273.
Impellizzeri, F. M., Rampinini, E. and Marcora, S. M. (2005) 'Physiological assessment of aerobic training in soccer', *Journal of Sports Sciences,* 23(6), pp. 583-592.
International Hockey Federation (2023a) *FIH Rankings: Outdoor*. Available at: https://www.fih.hockey/outdoor-hockey-rankings (Accessed: Aug. 26, 2023.
International Hockey Federation (2023b) *Our Members*. Available at: https://www.fih.hockey/about-fih/ourmembers (Accessed: Aug. 26, 2023.

Jackson, B. M., Polglaze, T., Dawson, B., King, T. and Peeling, P. (2018) 'Comparing global positioning system and global navigation satellite system measures of team-sport movements', *International Journal of Sports Physiology and Performance,* 13(8), pp. 1005-1010.

Jean-Christophe, H., Philippe, N., Michel, S., Jean-François, T. and François, D. (2018) 'Effects of intensity distribution changes on performance and on training loads quantification', *Biology of Sport,* 35(1), pp. 67–74.

Jeffries, A. C., Wallace, L., Coutts, A. J., McLaren, S. J., McCall, A. and Impellizzeri, F. M. (2020) 'Athlete-reported outcome measures for monitoring training responses: A systematic review of risk of bias and measurement property quality according to the COSMIN guidelines', *International Journal of Sports Physiology and Performance,* 15(9), pp. 1203-1215.

Jennings, D., Aughey, R. J., Cormack, S. J. and Coutts, A. J. (2012a) 'GPS analysis of an international field hockey tournament', *International Journal of Sports Physiology and Performance,* 7(3), pp. 224-231.

Jennings, D., Boyd, L., Aughey, R. J., Cormack, S. and Coutts, A. J. (2010a) 'The validity and reliability of GPS units for measuring distance in team sport specific running patterns', *International Journal of Sports Physiology and Performance,* 5(3), pp. 328-341.

Jennings, D., Cormack, S., Coutts, A. J., Boyd, L. J. and Aughey, R. J. (2010b) 'Variability of GPS units for measuring distance in team sport movements', *International Journal of Sports Physiology and Performance,* 5(4), pp. 565-569.

Jennings, D., Cormack, S. J., Coutts, A. J. and Aughey, R. J. (2012b) 'GPS analysis of an international field hockey tournament', *International Journal of Sports Physiology and Performance,* 7(3), pp. 224-231.

Jennings, D. H., Cormack, S. J., Coutts, A. J. and Aughey, R. J. (2012c) 'International field hockey players perform more high-speed running than national-level counterparts', *Journal of Strength and Conditioning Research,* 26(4), pp. 947-952.

Johnson, U.,診ttä, G., Ivarsson, A., Alvmyren, I. and Karlsson, M. (2016) 'An ultra- runner's experience of physical and emotional challenges during a 10- week continental run', *International Journal of Sport and Exercise Psychology,* 14(1), pp. 72-84.

Johnston, R. D., Gibson, N. V., Twist, C., Gabbett, T. J., Macnay, S. A. and Macfarlane, N. G. (2013) 'Physiological responses to an intensified period of rugby league competition', *Journal of Strength and Conditioning Research,* 27(3), pp. 643-654.

Johnston, R. J., Watsford, M. L., Austin, D. J., Pine, M. J. and Spurrs, R. W. (2015) 'An examination of the relationship between movement demands and rating of perceived exertion in Australian footballers', *The Journal of Strength and Conditioning Research,* 29(7), pp. 2026-2033.

Johnston, R. J., Watsford, M. L., Kelly, S. J., Pine, M. J. and Spurrs, R. W. (2014) 'Validity and interunit reliability of 10 Hz and 15 Hz GPS units for assessing athlete movement demands', *Journal of Strength and Conditioning research,* 28(6), pp. 1649-1655.

Johnston, R. J., Watsford, M. L., Pine, M. J., Spurrs, R. W., Murphy, A. J. and Pruyn, E. C. (2012) 'The validity and reliability of 5-Hz global positioning system units to measure team sport movement demands', *Journal of Strength and Conditioning Research,* 26(3), pp. 758-765.

Jones, A. (1991) 'Test and Measurement: 300-yard Shuttle Run', *Strength & Conditioning Journal,* 13(2), pp. 56-60.

Jones, A. M. and Doust, J. H. (1996) 'A 1% treadmill grade most accurately reflects the energetic cost of outdoor running', *Journal of Sports Sciences,* 14(4), pp. 321-327.

Jones, C. M., Griffiths, P. C. and Mellalieu, S. D. (2017) 'Training load and fatigue marker associations with injury and illness: A systematic review of longitudinal studies', *Sports Medicine,* 47(5), pp. 943-974.

Jonge, X. (2003) 'Effects of the menstrual cycle on exercise performance', *Sports Medicine,* 33(11), pp. 833-851.

Jurimae, J., Maestu, J., Purge, P., Jurimae, T. and Soot, T. (2002) 'Relations among heavy training stress, mood state, and performance for male junior rowers', *Perceptual and Motor Skills,* 95(2), pp. 520-526.

Kallus, W. and Kellmann, M. (2016) *The Recovery-Stress Questionnaires: User Manual.*

Kellmann, M. 2010. Preventing overtraining in athletes in high-intensity sports and stress/ recovery monitoring. Oxford, UK.

Kellmann, M., Altenburg, D., Lormes, W. and Steinacker, J. M. (2001) 'Assessing stress and recovery during preparation for the world championships in rowing', *The Sport Psychologist,* 15(2), pp. 151-167.

Kellmann, M. and Kallus, K. W. (2001) *Recovery-stress questionnaire for athletes : user manual.* Champaign, IL: Human Kinetics.

Kellmann, M. and Klaus-Dietrich, G. (2000) 'Changes in stress and recovery in elite rowers during preparation for the Olympic Games', *Medicine and Science in Sports and Exercise,* 32(3), pp. 676-683.

Kelly, D. M., Strudwick, A. J., Atkinson, G., Drust, B. and Gregson, W. (2020) 'Quantification of training and match-load distribution across a season in elite English Premier League soccer players', *Sci Med Football,* 4(1), pp. 59-67.

Kentta, G., Hassmen, P. and Raglin, J. S. (2001) 'Training practices and overtraining syndrome in Swedish age-group athletes', *International Journal of Sports Medicine,* 22(6), pp. 460-465.

Kevin, A. M. and James, R. B. (2015) 'Less Is more: The physiological basis for tapering in endurance, strength, and power athletes', *Sports,* 3(3), pp. 209-218.

Kim, T., Cha, J. H. and Park, J. C. (2018) 'Association between in-game performance parameters recorded via global positioning system and sports injuries to the lower extremities in elite female field hockey players', *Cluster Computing-the Journal of Networks Software Tools and Applications,* 21(1), pp. 1069-1078.

Kölling, S., Hitzschke, B., Holst, T., Ferrauti, A., Meyer, T., Pfeiffer, M. and Kellmann, M. (2015) 'Validity of the Acute Recovery and Stress Scale: Training monitoring of the German junior national field hockey team', *International Journal of Sports Science & Coaching,* 10(2-3), pp. 529-542.

Konarski, J., Matuszynski, M. and Strzelczyk, R. (2006) 'Different team defense tactics and heart rate during a field hockey match', *Studies in Physical Culture & Tourism,* 13, pp. 145-147.

Konerth, N. M. 2019. Unpublished Work - A Comparison and Analysis of Internal and External Training Load Measures in Female Hockey Athletes. Durham University.

Kraft, J. A., Laurent, M. L., Green, J. M., Helm, J., Roberts, C. and Holt, S. (2018) 'Examination of coach and player perceptions of recovery and exertion', *Journal of Strength and Conditioning Research,* 34(5), pp. 1383-1391.

Krueger, M., Costello, J. T., Stenzel, M., Mester, J. and Wahl, P. (2019) 'The physiological effects of daily cold-water immersion on 5-day tournament performance in international standard youth field-hockey players', *European Journal of Applied Physiology,* 120, pp. 295-305.

Laffite, L. P., Mille-Hamard, L., Koralsztein, J. P. and Billat, V. L. (2003) 'The effects of interval training on oxygen pulse and performance in supra-threshold runs', *Archives of Physiology and Biochemistry,* 111(3), pp. 202-210.

Lakens, D. (2017) 'Equivalence tests: A practical primer for t tests, correlations, and meta-analyses', *Social Psychological & Personality Science,* 8(4), pp. 355-362.

Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M. and Dienes, Z. (2020) 'Improving Inferences About Null Effects With Bayes Factors and Equivalence Tests', *The Journals of Gerontology: Series B,* 75(1), pp. 45-57.

Lakens, D., Scheel, A. M. and Isager, P. M. (2018) 'Equivalence testing for psychological research: A tutorial', *Advances in Methods and Practices in Psychological Science,* 1(2), pp. 259-269.

Lambert, M. and Borresen, J. (2006) 'A theoretical basis of monitoring fatigue: A practical approach for coaches', *International Journal of Sports Science & Coaching,* 1(4), pp. 371-388.

Leunes, A. and Burger, J. (2000) 'Profile of mood states research in sport and exercise psychology: past, present, and future', *Journal of Applied Sport Psychology,* 12(1), pp. 5-15.

Liu, H., Zhao, G., Gómez, M. A., Molinuevo, J. S., Giménez, J. V. and Kang, H. (2013) 'Time-motion analysis on Chinese male field hockey players', *International Journal of Performance Analysis in Sport,* 13(2), pp. 340-352.

López-Fernández, J., Gallardo, L., Fernández-Luna, Á., Villacañas, V., García-Unanue, J. and Sánchez-Sánchez, J. (2019) 'Pitch size and game surface in different small-sided games. Global indicators, activity profile, and acceleration of female soccer players', *Journal of strength and conditioning research,* 33(3), pp. 831-838.

Lucia, A., Hoyos J Fau - Santalla, A., Santalla A Fau - Earnest, C., Earnest C Fau - Chicharro, J. L. and Chicharro, J. L. (2003) 'Tour de France versus Vuelta a España: which is harder?', *Medicine and Science in Sports and Exercise,* 35(5).

Lundqvist, C. and Kenttä, G. (2010) 'Positive emotions are not simply the absence of the negative ones: Development and validation of the Emotional Recovery Questionnaire (EmRecQ)', *Sport Psychologist,* 24(4), pp. 468-488.

Lythe, J. (2008) *The physical demands of elite men's field hockey and the effects of differing substitution methods on the physical and technical outputs of strikers during match play.* Masters of Health Science, Auckland University of Technology.

Lythe, J. and Kilding, A. E. (2011) 'Physical demands and physiological responses during elite field hockey', *International Journal of Sports Medicine,* 32(7), pp. 523-528.

Lythe, J. and Kilding, A. E. (2013) 'The effect of substitution frequency on the physical and technical outputs of strikers during field hockey match play', *International Journal of Performance Analysis in Sport,* 13(3), pp. 848-859.

Macleod, H., Morris, J., Nevill, A. and Sunderland, C. (2009) 'The validity of a non-differential global positioning system for assessing player movement patterns in field hockey', *Journal of Sports Sciences,* 27(2), pp. 121-128.

Macutkiewicz, D. and Sunderland, C. (2011) 'The use of GPS to evaluate activity profiles of elite women hockey players during match-play', *Journal of Sports Sciences,* 29(9), pp. 967-973.

Malone, J. J., Lovell, R., Varley, M. C. and Coutts, A. J. (2017) 'Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport', *International Journal of Sports Physiology and Performance,* 12(Suppl 2), pp. S218-S226.

Malone, S. and Collins, K. (2016) 'Relationship between individualized training impulse and aerobic fitness measures in hurling players across a training period.', *Journal of Strength and Conditioning Research,* 30(11), pp. 3140-3145.

Malone, S., Hughes, B., Collins, K. and Akubat, I. (2018) 'Methods of monitoring training load and their association with changes across fitness measures in hurling players', *Journal of Strength and Conditioning Research,* 34(1), pp. 225-234.

Manzi, V., Bovenzi, A., Franco Impellizzeri, M., Carminati, I. and Castagna, C. (2013) 'Individual training- load and aerobic- fitness variables in premiership soccer players during the precompetitive season', *Journal of Strength and Conditioning Research,* 27(3), pp. 631-636.

Manzi, V., Iellamo, F., Impellizzeri, F., D'ottavio, S. and Castagna, C. (2009) 'Relation between individualized training impulses and performance in distance runners', *Medicine & Science in Sports & Exercise,* 41(11), pp. 2090-2096.

Mara, J. K., Thompson, K. G., Pumpa, K. L. and Ball, N. B. (2015) 'Periodization and physical performance in elite female soccer players', *International Journal of Sports Physiology and Performance,* 10(5), pp. 664-669.

Martinent, G., Decret, J.-C., Filaire, E., Isoard-Gautheur, S. and Ferrand, C. (2014) 'Evaluations of the psychometric properties of the recovery-stress questionnaire for athletes among a sample of young French table tennis players', *Psychological Reports,* 114(2), pp. 326-340.

Maupin, D., Schram, B., Canetti, E. and Orr, R. (2020) 'The relationship between acute: chronic workload ratios and injury risk in sports: A systematic review', *Open Access Journal of Sports Medicine,* 11, pp. 51-75.

McCorry, L. K. (2007) 'Physiology of the autonomic nervous system', *American Journal of Pharmaceutical Education,* 71(4), pp. 78-78.

McGuinness, A., Malone, S., Hughes, B., Collins, K. and Passmore, D. (2019) 'Physical activity and physiological profiles of elite international female field hockey players across the quarters of competitive match play', *Journal of Strength and Conditioning Research,* 33(9), pp. 2513-2522.

McGuinness, A., Malone, S., Petrakos, G. and Collins, K. (2017) 'The physical and physiological demands of elite international female field hockey players during competitive match-play', *Journal of Strength and Conditioning Research,* 33(11), pp. 3105-3113.

McGuinness, A., McMahon, G., Malone, S., Kenna, D., Passmore, D. and Collins, K. (2018) 'Monitoring wellness, training load, and running performance during a major international female field hockey tournament', *Journal of Strength and Conditioning Research,* 34(8), pp. 2312-2320.

McKeon, J. M. M. and McKeon, P. O. (2012) 'Evaluation of joint position recognition measurement variables associated with chronic ankle instability: A meta-analysis', *Journal of Athletic Training,* 47(4), pp. 444-456.

McLaren, S., Macpherson, T., Coutts, A., Hurst, C., Spears, I. and Weston, M. (2018) 'The relationships between internal and external measures of training load and intensity in team sports: A meta-analysis', *Sports Medicine,* 48(3), pp. 641-658.

McLaren, S. J., Smith, A., Spears, I. R. and Weston, M. (2017) 'A detailed quantification of differential ratings of perceived exertion during team-sport training', *Journal of Science and Medicine in Sport,* 20(3), pp. 290-295.

McMahon, G. E. and Kennedy, R. A. (2019) 'Changes in player activity profiles after the 2015 FIH rule changes in elite women's hockey', *Journal of Strength & Conditioning Research* 33(11), pp. 3114-3122.

McMahon, G. E., Sharp, L.-A. and Kennedy, R. A. (2021) 'Running performance Is correlated with creatine kinase levels and muscle soreness during an Olympic games in hockey', *International Journal of Sports Physiology and Performance,* 16(10), pp. 1393-1400.

McNulty, K. L., Ansdell, P., Goodall, S., Thomas, K., Elliott-Sale, K. J., Howatson, G. and Hicks, K. M. (2023) 'The symptoms experienced by naturally menstruating women and oral contraceptive pill users and their perceived effects on exercise performance and recovery time posttraining', *Women in Sport and Physical Activity Journal*, pp. 1-13.

Meeusen, R., Duclos, M., Foster, C., Fry, A., Gleeson, M., Nieman, D., Raglin, J., Rietjens, G., Steinacker, J. and Urhausen, A. (2013) 'Prevention, diagnosis, and treatment of the overtraining syndrome: joint consensus statement of the European College of Sport Science and the American College of Sports Medicine', *Medicine & Science In Sports & Exercise,* 45(1), pp. 186-205.

Milanez, V. F., Ramos, S. P., Okuno, N. M., Boullosa, D. A. and Nakamura, F. Y. (2014) 'Evidence of a non-linear dose-response relationship between training load and stress markers in elite female futsal players', *Journal of Sports Science and Medicine* 13(1), pp. 22-29.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (2009) 'Preferred reporting items for systematic reviews and meta- analyses: the PRISMA statement', *Annals of Internal Medicine* 151(4), pp. 264-269.

Morencos, E., Casamichana, D., Torres, L., Romero-Moraleda, B., Haro, X. and Rodas, G. (2019) 'Kinematic demands of international competition in women's field hockey', *Apunts Educacion Fisica Y Deportes,* (137), pp. 56-70.

Morencos, E., Romero-Moraleda, B., Castagna, C. and Casamichana, D. (2018) 'Positional comparisons in the impact of fatigue on movement patterns in hockey', *International Journal of Sports Physiology and Performance,* 13(9), pp. 1149-1157.

Moreno-Pérez, V., Sánchez-Migallón, V., Domínguez, R., Fernández-Elías, V., Fernández-Fernández, J., Pérez-López, A. and López-Samanes, A. (2019) 'The acute effect of match-play on hip range of motion and isometric strength in elite tennis players', *PeerJ,* 7(11), pp. e7940.

Morgan, W. P., Brown, D. R., Raglin, J. S., Connor, P. J. and Ellickson, K. A. (1987a) 'Psychological monitoring of overtraining and staleness', *British Journal of Sports Medicine,* 21(3), pp. 107-114.

Morgan, W. P., Costill, D. L., Flynn, M. G., Raglin, J. S. and O'Connor, P. J. (1988) 'Mood disturbance following increased training in swimmers', *Medicine & Science In Sports & Exercise,* 20(4), pp. 408-14.

Morgan, W. P., O'Connor, P. J., Sparling, P. B. and Pate, R. R. (1987b) 'Psychological characterization of the elite female distance runner', *International Journal of Sports Medicine,* 8(Suppl 2), pp. 124-31.

Morgan, W. P., O'Connor, P. J., Ellickson, K. A., & Bradley, P. W. (1988) 'Personality structure, mood states, and performance in elite male distance runners', *International Journal of Sport Psychology,* 19(4), pp. 247-263.

Morton, R. H., Fitz-Clarke, J. R. and Banister, E. W. (1990) 'Modeling human performance in running', *Journal of Applied Physiology,* 69(3), pp. 1171.

Mujika, I. (1998) 'The influence of training characteristics and tapering on the adaptation in highly trained individuals: a review', *International Journal of Sports Medicine,* 19(7), pp. 439-446.

Murray, N. B., Gabbett, T. J., Townshend, A. D. and Blanch, P. (2017) 'Calculating acute:chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages', *British Journal of Sports Medicine,* 51(9), pp. 749-754.

Nässi, A., Ferrauti, A., Meyer, T., Pfeiffer, M. and Kellmann, M. (2017a) 'Development of two short measures for recovery and stress in sport', *European Journal of Sport Science,* 17(7), pp. 894-903.

Nässi, A., Ferrauti, A., Meyer, T., Pfeiffer, M. and Kellmann, M. (2017b) 'Psychological tools used for monitoring training responses of athletes', *Performance Enhancement & Health,* 5(4), pp. 125-133.

Nelson, M. J., Thomson, R. L., Rogers, D. K., Howe, P. R. C. and Buckley, J. D. (2014) 'Maximal rate of increase in heart rate during the rest- exercise transition tracks reductions in exercise performance when training load is increased', *Journal of Science and Medicine in Sport,* 17(1), pp. 129-133.

NHS Digital, Services, N.H. (2017) *Sexual and Reproductive Health Services - England, 2016/17*.

Nicolas, M., Vacher, P., Martinent, G. and Mourot, L. (2019) 'Monitoring stress and recovery states: Structural and external stages of the short version of the RESTQ sport in elite swimmers before championships', *Journal of Sport and Health Science,* 8(1), pp. 77-88.

Noblett, H., Hudson, S., Killey, J. and Fish, M. (2023) 'The physical and physiological match-play locomotor activity profiles of elite domestic male field hockey', *Journal of Sports Science & Medicine,* 22(2), pp. 273.

Nunes, J. A., Moreira, A., Crewther, B. T., Nosaka, K., Viveiros, L. and Aoki, M. S. (2014) 'Monitoring training load, recovery-stress state, immune-endocrine responses, and physical performance in elite female basketball players during a periodized training program', *Journal of Strength and Conditioning Research,* 28(10), pp. 2973-2980.

Nygaard Falch, H., Guldteig Rædergård, H. and van den Tillaar, R. (2019) 'Effect of different physical training forms on change of direction ability: A systematic review and meta-analysis', *Sports Medicine - Open,* 5(1), pp. 1-37.

O'Connor, P. J., Raglin, J. S. and Morgan, W. P. (1996) 'Psychometric correlates of perception during arm ergometry in males and females', *International Journal of Sports Medicine,* 17(6), pp. 462-466.

Oliveira-Silva, I. A.-O., Silva, V. A., Cunha, R. M. and Foster, C. (2018) 'Autonomic changes induced by pre-competitive stress in cyclists in relation to physical fitness and anxiety', *Plos One,* 13(12), pp. e0209834.

Olympic.org (2023) *Hockey*. Available at: https://www.olympic.org/hockey.
Oosthuyse, T. and Bosch, A. N. (2010) 'The effect of the menstrual cycle on exercise metabolism: implications for exercise performance in eumenorrhoeic women', *Sports medicine,* 40(3), pp. 207-227.
Padulo, J., Powell D Fau - Milia, R., Milia R Fau - Ardigò, L. P. and Ardigò, L. P. (2013) 'A paradigm of uphill running', *Plos One,* 8(7), pp. e69006.
Parrado, E., Cervantes, J., Pintanel, M., Rodas, G. and Capdevila, L. (2010) 'Perceived tiredness and heart rate variability in relation to overload during a field hockey world cup', *Perceptual and Motor Skills,* 110(3), pp. 699-713.
Passfield, L., Murias, J. M., Sacchetti, M. and Nicolò, A. (2022) 'Validity of the training-load concept', *International Journal of Sports Physiology and Performance,* 17(4), pp. 507-514.
Pate, R. R. (1988) 'The evolving definition of physical fitness', *Quest,* 40(3), pp. 174-179.
Pearson, D., Faigenbaum, A., Conley, M. and Kraemer, W. (2000) 'The National Strength and Conditioning Association's basic guidelines for the resistance training of athletes', *Strength and Conditioning Journal,* 22(4), pp. 14.
Perrotta, A. S., Held, N. J. and Warburton, D. E. R. (2017) 'Examination of internal training load parameters during the selection, preparation and competition phases of a mesocycle in elite field hockey players', *International Journal of Performance Analysis in Sport,* 17(5), pp. 813-821.
Perrotta, A. S., Koehle, M. S., White, M. D., Taunton, J. E. and Warburton, D. E. R. (2019a) 'Consecutive non-training days over a weekend for assessing cardiac parasympathetic variation in response to accumulated exercise stress', *European Journal of Sport Science,* 20(8), pp. 1072-1082.
Perrotta, A. S., Taunton, J. E., Koehle, M. S., White, M. D. and Warburton, D. E. R. (2019b) 'Monitoring the prescribed and experienced heart rate-derived training loads in elite field hockey players', *Journal of Strength and Conditioning Research,* 33(5), pp. 1394-1399.
Perrotta, A. S. and Warburton, D. E. R. (2018) 'A comparison of sessional ratings of perceived exertion to cardiovascular indices of exercise intensity during competition in elite field hockey players', *Biomedical Human Kinetics,* 10(1), pp. 157-162.
Phillips, E., Davids, K., Renshaw, I. and Portus, M. (2010) 'Expert performance in sport and the dynamics of talent development', *Sports Medicine,* 40(4), pp. 271-283.
Pinheiro, J., Bates, D., DebRoy, S. and Sarkar, D. (2021) 'nlme : Linear and Nonlinear Mixed Effects Models'. Available at: https://CRAN.R-project.org/package=nlme.
Pisk, J. (2014) 'Sport science: ontological and methodological considerations', *Physical Culture and Sport,* 61(1), pp. 5.
Podgórski, T. and Pawlak, M. (2011) 'A half century of scientific research in field hockey', *Human Movement,* 12(2), pp. 108-123.
Polglaze, T., Dawson, B., Buttfield, A. and Peeling, P. (2018) 'Metabolic power and energy expenditure in an international men's hockey tournament', *Journal of Sports Sciences,* 36(2), pp. 140-148.
Polglaze, T., Dawson, B., Hiscock, D. J. and Peeling, P. (2015) 'A comparative analysis of accelerometer and time - motion data in elite men's hockey training and competition', *International Journal of Sports Physiology and Performance,* 10(4), pp. 446-451.

Pope, C. C., Penney, D. and Smith, T. B. (2018) 'Overtraining and the complexities of coaches & decision-making: managing elite athletes on the training cusp', *Reflective Practice,* 19(2), pp. 145-166.

Portas, M. D., Harley, J. A., Barnes, C. A. and Rush, C. J. (2010) 'The Validity and Reliability of 1-Hz and 5-Hz Global Positioning Systems for Linear, Multidirectional, and Soccer-Specific Activities', *International Journal of Sports Physiology and Performance,* 5(4), pp. 448-458.

Pueo, B., Jimenez-Olmedo, J., Penichet-Tomas, A. and Espina Agullo, J. (2017) 'Analysis of time-motion and heart rate in elite male and female beach handball', *Journal of Sports Science and Medicine,* 16(4), pp. 450-458.

Raeder, C., Wiewelhove, T., De Paula Simola, R. A., Kellmann, M., Meyer, T., Pfeiffer, M. and Ferrauti, A. (2016) 'Assessment of fatigue and recovery in male and female athletes after 6 days of intensified strength training', *Journal of Strength and Conditioning Research,* 30(12), pp. 3412-3427.

Raglin, J. S. and Morgan, W. P. (1994) 'Development of a scale for use in monitoring training-induced distress in athletes', *International Journal of Sports Medicine,* 15(2), pp. 84-88.

Raglin, J. S., Morgan, W. P. and Luchsinger, A. E. (1990) 'Mood and self-motivation in successful and unsuccessful female rowers', *Medicine & Science In Sports & Exercise,* 22(6), pp. 849-853.

Ravé, G., Granacher, U., Boullosa, D., Hackney, A. C. and Zouhal, H. (2020) 'How to use global positioning systems (GPS) data to monitor training load in the “real world” of elite soccer', *Frontiers in Physiology,* 11, pp. 944.

Rawstorn, J. C., Maddison, R., Ali, A., Foskett, A. and Gant, N. (2014) 'Rapid directional change degrades GPS distance measurement validity during intermittent intensity running', *Plos One,* 9(4), pp. e93693.

Reddy, P., Dias, I., Holland, C., Campbell, N., Nagar, I., Connolly, L., Krustrup, P. and Hubball, H. (2017) 'Walking football as sustainable exercise for older adults - A pilot investigation', *European Journal of Sport Science,* 17(5), pp. 638-645.

Reilly, T. and Borrie, A. (1992) 'Physiology applied to field hockey', *Sports Medicine,* 14(1), pp. 10-26.

Reilly, T. and Seaton, A. (1990) 'Physiological strain unique to field hockey', *The Journal of Sports Medicine and Physical Fitness,* 30(2), pp. 142-146.

Robertson, S., Bartlett, J. D. and Gastin, P. B. (2017) 'Red, amber, or green? Athlete monitoring in team sport: The need for decision-support systems', *International Journal of Sports Physiology and Performance,* 12(S2), pp. S2-73-79.

Robson-Ansley, P., Blannin, A. and Gleeson, M. (2007) 'Elevated plasma interleukin- 6 levels in trained male triathletes following an acute period of intense interval training', *European Journal of Applied Physiology,* 99(4), pp. 353-360.

Robson-Ansley, P. J., Gleeson, M. and Ansley, L. (2009) 'Fatigue management in the preparation of Olympic athletes', *Journal of Sports Sciences,* 27(13), pp. 1409-1420.

Ross, J., Hecksteden, A., Fullagar, H. and Meyer, T. (2017) 'The effects of menstrual cycle phase on physical performance in female soccer players', *PLoS One,* 12(3), pp. e0173951.

Rousanoglou, E. N., Georgiadis, G. V. and Boudolos, K. D. (2008) 'Muscular strength and jumping performance relationships in young women athletes', *Journal of Strength and Conditioning Research,* 22(4), pp. 1375-1378.

Rushall, B. S. (1990) 'A tool for measuring stress tolerance in elite athletes', *Journal of Applied Sport Psychology,* 2(1), pp. 51-66.

Sampson, J. A., Fullagar, H. H. K. and Murray, A. (2017) 'Evidence is needed to determine if there is a better way to determine the acute:chronic workload', *British Journal of Sports Medicine,* 51(7), pp. 621-622.

Sanders, D., Abt, G., Hesselink, M. K. C., Myers, T. and Akubat, I. (2017) 'Methods of monitoring training load and their relationships to changes in fitness and performance in competitive road cyclists', *International Journal of Sports Physiology and Performance,* 12(5), pp. 668-675.

Saw, A. E., Main, L. C. and Gastin, P. B. (2016) 'Monitoring the athlete training response: subjective self- reported measures trump commonly used objective measures: A systematic review', *British Journal of Sports Medicine,* 50(5), pp. 281-291.

Schneider, C., Hanakam, F., Wiewelhove, T., Daweling, A., Kellmann, M., Meyer, T., Pfeiffer, M. and Ferrauti, A. (2018) 'Heart rate monitoring in team sports-a conceptual framework for contextualizing heart rate measures for training and recovery prescription', *Frontiers in Physiology,* 9, pp. 639.

Scotland, J. (2012) 'Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms', *English Language Teaching,* 5(9), pp. 9-16.

Scott, B. R., Lockie, R. G., Knight, T. J., Clark, A. C. and De Jonge, X. A. K. J. (2013) 'A comparison of methods to quantify the in- season training load of professional soccer players', *International Journal of Sports Physiology and Performance,* 8(2), pp. 195-202.

Scott, M., Scott, T. and Kelly, V. (2016) 'The validity and reliability of global positioning systems in team sport: A brief review', *Journal of Strength and Conditioning Research,* 30(5), pp. 1470-1490.

Scott, T., Black, C., Quinn, J. and Coutts, A. (2012) 'Validity and reliability of the session-RPE ,ethod for quantifying training in Australian football: A comparison of the CR10 and CR100 scales', *Journal of Strength and Conditioning Research,* 27(1), pp. 270-276.

Sell, K. M. and Ledesma, A. B. (2016) 'Heart rate and energy expenditure in Division I field hockey players during competitive play', *Journal of Strength and Conditioning Research,* 30(8), pp. 2122-2128.

Shushan, T., Lovell, R., Buchheit, M., Scott, T., Norris, D. and McLaren, S. (2022) 'Submaximal fitness test in team sports: A systematic review and meta-analysis of exercise heart rate measurement properties', *Sports Medicine - Open,* 9(1), pp. 21.

Singh, T. K. R., Guelfi, K. J., Landers, G., Dawson, B. and Bishop, D. (2011) 'A comparison of muscle damage, soreness and performance following a simulated contact and non-contact team sport activity circuit', *Journal of Science and Medicine in Sport,* 14(5), pp. 441-446.

Souglis, A., Bogdanis, G. C., Giannopoulou, I., Papadopoulos, C. and Apostolidis, N. (2015a) 'Comparison of inflammatory responses and muscle damage indices following a soccer, basketball, volleyball and handball game at an elite competitive level', *Research in Sports Medicine,* 23(1), pp. 59-72.

Souglis, A. G., Papapanagiotou, A., Bogdanis, G. C., Travlos, A. K., Apostolidis, N. G. and Geladas, N. D. (2015b) 'Comparison of inflammatory responses to a soccer match between elite male and female players', *Journal of strength and conditioning research,* 29(5), pp. 1227-1233.

Sperlich, B. and Holmberg, H.-C. (2017) 'The responses of elite athletes to exercise: An all-day, 24-h integrative view Is required!', *Frontiers in Physiology,* 8, pp. 564.

Stagno, K. M., Thatcher, R. and Van Someren, K. A. (2007) 'A modified TRIMP to quantify the in-season training load of team sport players', *Journal of Sports Sciences,* 25(6), pp. 629-634.

Stanković, M., Gušić, M., Nikolić, S., Barišić, V., Krakan, I., Sporiš, G., Mikulić, I. and Trajković, N. (2021) '30-15 intermittent fitness test: A systematic review of studies, examining the VO2max estimation and training programming', *Applied Sciences,* 11(24), pp. 11792.

Starling, L. T., Nellemann, S., Parkes, A. and Lambert, M. I. (2019) 'The Fatigue and Fitness Test for Teams (FFITT): A practical option for monitoring athletes in a team as individuals', *European Journal of Sport Science,* 20(2), pp. 106-114.

Sunderland, C. D. and Edwards, P. L. (2017) 'Activity profile and between-match variation in elite male field hockey', *Journal of Strength and Conditioning Research,* 31(3), pp. 758-764.

Tawa, N. and Louw, Q. (2018) 'Biomechanical factors associated with running economy and performance of elite Kenyan distance runners: A systematic review', *Journal of Bodywork & Movement Therapies,* 22(1), pp. 1-10.

Taylor, J., Wright, A., Dischiavi, S., Townsend, M. and Marmon, A. (2017) 'Activity demands during multi-directional team sports: A systematic review', *Sports Medicine,* 47(12), pp. 2533-2551.

Taylor, R., Myers, T. D., Sanders, D., Ellis, M. and Akubat, I. (2021) 'The relationship between training load measures and next-day well-being in rugby union players', *Applied Sciences,* 11(13), pp. 5926.

Thomas, J. R. (2011) *Research Methods in Physical Activity.* 6th ed. edn. Champaign, IL: Champaign, IL : Human Kinetics.

Thomson, R. L., Bellenger, C. R., Howe, P. R. C., Karavirta, L. and Buckley, J. D. (2016a) 'Improved heart rate recovery despite reduced exercise performance following heavy training: A within-subject analysis', *Journal of Science and Medicine in Sport,* 19(3), pp. 255-259.

Thomson, R. L., Rogers, D. K., Howe, P. R. C. and Buckley, J. D. (2016b) 'Effect of acute exercise- induced fatigue on maximal rate of heart rate increase during submaximal cycling', *Research in Sports Medicine,* 24(1), pp. 1-15.

Thornton, H. R., Delaney, J. A., Duthie, G. M. and Dascombe, B. J. (2019) 'Developing athlete monitoring systems in team sports: Data analysis and visualization', *International Journal of Sports Physiology and Performance,* 14(6), pp. 698-705.

Tim, J. G., George, P. N., Eric, O., Johan, P., Nick, J., Daniel, M., Gil, R., Tom, M., Dan, H., Adam, B. and Allan, R. (2017) 'The athlete monitoring cycle: a practical guide to interpreting and applying training monitoring data', *British Journal of Sports Medicine,* 51(20), pp. 1451.

Tobar, D. A. (2005) 'Overtraining and staleness: The importance of psychological monitoring', *International Journal of Sport and Exercise Psychology,* 3(4), pp. 455-468.

Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A. and Clubb, J. (2022) 'Tracking systems in team sports: A narrative review of applications of the data and sport specific analysis', *Sports Medicine - Open,* 8(1), pp. 1-22.

Tromp, M. and Holmes, L. (2011) 'The effect of free-hit rule changes on match variables and patterns of play in international standard women's field hockey', *International Journal of Performance Analysis in Sport,* 11(2), pp. 376-391.

Tsampoukos, A., Peckham, E., James, R. and Nevill, M. (2010) 'Effect of menstrual cycle phase on sprinting performance', *European Journal of Applied Physiology,* 109(4), pp. 659-667.

Tuft, K. and Kavaliauskas, M. (2020) 'Relationship between internal and external training load in field hockey', *International Journal of Strength and Conditioning,* 1(1), pp. 1-11.

Urhausen, A. and Kindermann, W. (2002) 'Diagnosis of overtraining', *Sports Medicine,* 32(2), pp. 95-102.

Uusitalo, A. L. (2001) 'Overtraining: making a difficult diagnosis and implementing targeted treatment', *The Physician and Sportsmedicine,* 29(5), pp. 35-50.

Valladares-Rodríguez, S., Rey, E., Mecías-Calvo, M., Barcala-Furelos, R. and Bores-Cerezal Antonio, J. (2017) 'Reliability and usefulness of the 30-15 intermittent fitness test in male and female professional futsal players', *Journal of Human Kinetics,* 60(1), pp. 191-198.

Van Hooren, B., Fuller, J. T., Buckley, J. D., Miller, J. R., Sewell, K., Rao, G., Barton, C., Bishop, C. and Willy, R. W. (2020) 'Is motorized treadmill running biomechanically comparable to overground running? A systematic review and meta-analysis of cross-over studies', *Sports Medicine,* 50(4), pp. 785-813.

Vanrenterghem, J., Nedergaard, N. J., Robinson, M. A. and Drust, B. (2017) 'Training load monitoring in team sports: A novel framework separating physiological and biomechanical load-adaptation pathways', *Sports Medicine,* 47(11), pp. 2135-2142.

Verde, T., Thomas, S. and Shephard, R. J. (1992) 'Potential markers of heavy training in highly trained distance runners', *British Journal of Sports Medicine,* 26(3), pp. 167-175.

Vescovi, J. (2016) 'Locomotor, heart-rate, and metabolic power characteristics of youth women's field hockey: Female athletes in motion (FAiM) study', *Research Quarterly for Exercise and Sport,* 87(1), pp. 68-77.

Vescovi, J. D. (2014) 'Impact of maximum speed on sprint performance during high-level youth female field hockey matches: Female athletes in motion (FAiM) study', *International Journal of Sports Physiology and Performance,* 9(4), pp. 621-626.

Vescovi, J. D. (2019) 'Intra-individual variation of HRV during orthostatic challenge in elite male field hockey players', *Journal of Medical Systems,* 43(12), pp. 328.

Vescovi, J. D. and Frayne, D. H. (2015) 'Motion characteristics of division I college field hockey: Female athletes in motion (FAiM) study', *International Journal of Sports Physiology and Performance,* 10(4), pp. 476-481.

Vescovi, J. D. and Klas, A. (2018) 'Accounting for the warm-up: describing the proportion of total session demands in women's field hockey–Female Athletes in Motion (FAiM) study', *International Journal of Performance Analysis in Sport,* 18(5), pp. 868-880.

Vescovi, J. D., Klas, A. and Mandic, I. (2019) 'Investigating the relationships between load and recovery in women's field hockey - Female athletes in motion (FAiM) study', *International Journal of Performance Analysis in Sport,* 19(5), pp. 672-682.

Vinson, D. O. N., Gerrett, N. and James, D. V. B. (2018) 'Influences of playing position and quality of opposition on standardized relative distance covered in domestic women's field hockey: Implications for coaches', *Journal of Strength & Conditioning Research* 32(6), pp. 1770-1777.

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C. and Vandenbroucke, J. P. (2007) 'The strengthening the reporting of observational studies in Epidemiology

(STROBE) statement: Guidelines for reporting observational studies', *International Journal of Surgery,* 12(12), pp. 1495-1499.

Walker, A. J., McFadden, B. A., Sanders, D. J., Bozzini, B. N., Conway, S. P. and Arent, S. M. (2020) 'Early season hormonal and biochemical changes in division I field hockey players: Is fitness protective?', *Journal of Strength and Conditioning Research,* 34(4), pp. 975-981.

Wasserstein, R. L., Schirm, A. L. and Lazar, N. A. (2019) 'Moving to a world beyond "p < 0.05"', *The American Statistician,* 73(sup1), pp. 1-19.

Weaving, D., Jones, B., Till, K., Abt, G. and Beggs, C. (2017) 'The case for adopting a multivariate approach to optimize training Load quantification in team sports', *Frontiers in Physiology,* 8, pp. 1024.

Wehbe, G., Gabett, T. J., Dwyer, D., McLellan, C. and Coad, S. (2015) 'Monitoring neuromuscular fatigue in team-sport athletes using a cycle-ergometer test', *International Journal of Sports Physiology and Performance,* 10(3), pp. 292-297.

West, S. W., Clubb, J., Torres-Ronda, L., Howells, D., Leng, E., Vescovi, J. D., Carmody, S., Posthumus, M., Dalen-Lorentsen, T. and Windt, J. (2020) 'More than a metric: How training load is used in elite sport for athlete management', *International Journal of Sports Medicine,* 42(4), pp. 300-306.

White, A. D. and MacFarlane, N. (2013) 'Time-on-pitch or full-game gps analysis procedures for elite field hockey?', *International Journal of Sports Physiology and Performance,* 8(5), pp. 549-555.

White, A. D. and Macfarlane, N. G. (2015a) 'Analysis of international competition and training in men's field hockey by global positioning system and inertial sensor technology', *Journal of Strength and Conditioning Research,* 29(1), pp. 137-143.

White, A. D. and Macfarlane, N. G. (2015b) 'Contextual effects on activity profiles of domestic field hockey during competition and training', *Human Movement Science,* 40, pp. 422-431.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wiewelhove, T., Raeder, C., Meyer, T., Kellmann, M., Pfeiffer, M. and Ferrauti, A. (2015) 'Markers for routine assessment of fatigue and recovery in male and female team sport athletes during high-intensity interval training', *PLoS One,* 10(10), pp. e0139801.

Wikström-Frisén, L., Nordström, A., Mincheva-Nilsson, L. and Larsén, K. (2016) 'Impact of season and oral contraceptive use on cortisol levels in physically active women', *Journal Of Exercise, Sports & Orthopedics,* 3(2).

Wilson, J. (2019) *Preparing Vector Devices for a Session*. Available at: https://support.catapultsports.com/hc/en-us/articles/360000916315-Preparing-Vector-Devices-for-a-Session (Accessed: February 7 2020).

Yamamoto, H., Takemura, M., Iguchi, J., Tachibana, M., Tsujita, J. and Hojo, T. (2020) 'In-match physical demands on elite Japanese rugby union players using a global positioning system', *BMJ Open Sport Exerc Med,* 6(1), pp. 1-10.

Young, W., Hepner, J. and Robbins, D. (2012) 'Movement demands in australian rules football as indicators of muscle damage', *Journal of Strength and Conditioning Research,* 26(2), pp. 492-496.